Where are the Sophisticated Investors? Evidence from Separate Accounts

Valeria Fedyk*

[1]Department of Finance PhD, London Business School,

*To whom correspondence should be addressed; E-mail: vfedyk@london.edu.

November 2021

Abstract

This paper examines the top drivers of investor flows into US Separate Account Composites, whose investors purportedly represent some of the most sophisticated and affluent investors in the world. I find that for actively managed US Separate Account Composites, Morningstar rating is the most significant predictor of flows and supersedes more financially sophisticated metrics such as the CAPM model alpha, Fama-French 3 Factor model alpha, Fama-French-Carhart 4 Factor model alpha and other measures of risk-adjusted return. Surprisingly, the aforementioned results appear to hold for passively managed US Mutual Funds as well. With regards to performance, I find that while on average Separate Accounts outperformed the market and achieved positive alpha over the first 1991-2011 period, they failed to do so over the more recent 2012-2020 period, underscoring that these exclusive investment vehicles may not deliver consistent outperformance.

*I owe my sincere gratitude to Anna Pavlova at London Business School for our extensive discussions and her considerable support, advice and encouragement throughout the project. I would also like to thank Svetlana Bryzgalova, Roberto Cram, Joao Cocco, Marco Grotteria, Nicholas Hirschey, Narayan Naik, and Stephen Schaefer for helpful discussions and feedback. This paper was considerably enhanced as a result. Finally, I am grateful to seminar participants John Cochrane, Ken Singleton, and Lubos Pastor for their feedback on the paper in post-seminar discussions.

1 Introduction

In this paper, I assess which asset pricing model comes closest to the one that large and sophisticated accredited investors use by looking at the most significant drivers of separate account investor flows. Surprisingly, I find that even these purportedly most sophisticated investors tend to base their flows more on simple heuristics such as Morningstar ratings than on any of the more financially sophisticated measures of performance and ability such as factor model alphas.

My analysis centers on the US Separate Account Composites dataset from Morningstar, which is a novel dataset that has not been looked at in prior literature. This dataset deals with composites that include separate accounts of highly sophisticated investors such as pension funds, insurance companies, wealth management funds, and sovereign wealth funds. All of the investors into separate accounts are accredited investors, meaning there are no small retail investors who are generally taken to be less sophisticated than their institutional counterparts. In fact, the average minimum investment for these accounts is a hefty $10 million. Separate accounts also differ from pooled funds such as mutual funds because each account is managed and traded separately, and can include slight individual customization such as sin stock exclusions or tighter position constraints. Separate accounts thus offer investors direct ownership of the securities traded, and remove them from the risk of other investors trading in or out of the strategy.

I consider a variety of performance metrics, ranging from simple heuristics to multi-factor model alphas, as possible drivers of investor flows. These measures include the CAPM, Fama-French 3 Factor, and Fama-French-Carhart 4 Factor model alphas; market-adjusted return, benchmark-adjusted return, and excess of risk-free rate return; Morningstar ratings; and the Sharpe ratio and Information ratio. Out of these performance metrics, I find that the Morningstar rating has the highest significance in explaining investor flows for actively managed US separate account composites. In fact, an indicator variable of whether a fund has 5 stars agrees with whether or not the fund receives a positive inflow the following month nearly 61.47% of the time. The fraction of positive flows that a fund receives also increases monotonically as the number of Morningstar rating stars for the fund increases, ranging from an average of 27.83% for 1 star strategies to 60.60% for 5 star strategies. Thus the Morningstar rating appears to be a significant driver of flows for actively managed US separate account composites, and it appears to matter relatively more - both in terms of magnitude and significance - than other risk-adjusted return metrics or even the more financially sophisticated factor model alphas.

I also perform a direct pairwise comparison of the models considered using the flow sign test developed by Berk and van Binsbergen (2016). Reinforcing the illustrative evidence outlined above, this test shows that Morningstar rating predicts the sign of flows better than all other models considered at the 1% statistical significance level. While 3 Factor and 4 Factor models do perform slightly better than the CAPM in predicting flow sign, all of the factor models are statistically significantly worse at predicting flow sign than the Morningstar rating at the 1% significance level.

When I run a time fixed effects panel regression of flows on the various performance measures discussed using controls for past fund flows, fund size, and fund age and double clustering standard errors by fund and time, I also find that the Morningstar rating is a significant predictor of flows at the 1% level for separate account composites. In a comparison of the CAPM, 3 Factor, and 4 Factor alphas in driving flows, the 4 Factor alpha has the highest coefficient and significance. For the Sharpe ratio and Information ratio, I find that both are significant predictors of flows on their own but do not contribute much in combination with the other performance metrics.

I also look at a comparison of value style, growth style, and blend style mutual funds. I find no significant difference in the results between each style and the entire sample of mutual funds for the sign flow test. For each style, Morningstar ratings continue to be the best predictor of future flow sign, followed by the CAPM. When splitting the data sample over time, however, I do find a relative increase in the sophistication of models used: when comparing the first period (1991-2011) with the second period (2012-2019), the percentage of flow sign agreement decreases for all performance measures considered except for the 4 Factor alpha, whose percentage of flow sign agreement increases.

For passively managed mutual funds, flow sign continues to be predicted by prior month Morningstar rating and factor model alphas, though to a lesser degree than for the actively managed mutual funds. The Morningstar rating indicator of whether a fund has 5 stars or not now correctly predicts flow sign for the next month 62.5% of the time, while CAPM alpha sign agrees with next month flow sign 55% of the time. Since passively managed funds generally track the index without endeavouring to provide alpha, the fact that investor flows are nevertheless influenced by passive fund outperformance is a surprising finding.

A final contribution I make in this paper is by assessing the performance of the separate account composites. I find considerable variation in market-adjusted returns, benchmark-adjusted returns, and multi-factor model alphas over time. While the separate account composites show average gross and net of fees outperformance relative to the market in the first part of the sample period (1991-2011), the average relative outperformance turns negative, on both a gross and net basis, for the second part of the sample period (2012-2020). Due to the relative outperformance in one decade and underperformance in the following decade, I conclude that there is no conclusive evidence of consistent separate account outperformance.

The paper is organized as follows. In Section 2, I offer a review of the related literature. Section 3 describes the separate account composites and mutual fund datasets used and the methodology for calculating fund-level data points and fund factor loadings, alphas, and weighted alphas. In Section 4, I cover additional performance measures including the Morningstar rating, Sharpe ratio, and Information ratio. Section 5 presents the main results of the paper, with robustness checks included in Section 6. Section 7 presents possible theoretical explanations for my findings, and Section 8 concludes.

2 Literature Review

A central question in asset pricing has long been which asset pricing model is actually used by investors in the market. The celebrated Capital Asset Pricing Model (CAPM), first championed by William Sharpe and John Lintner in the mid-20th century and described in Sharpe (1964), was hailed as a landmark asset pricing model when it was first introduced. Taking into account both non-diversifiable systematic risk of an asset and the expected market and risk free returns, the model can provide the theoretical risk premium for any asset.

At the same time, the CAPM also makes certain simplifying assumptions that can preclude it from holding perfectly in the real world. For instance, it assumes a utility function which only takes into account the mean return and variance, rather than higher moments of risk. In addition, it is empirically difficult to identify exactly what the entire market portfolio should include - would it include only equity and fixed income investments, for instance, or also private equity and real estate holdings? What about human capital? Jagannathan and Wang (1996), for instance, relax the assumptions that the market portfolio is the value-weighted average of all stocks and that betas are constant over time. Instead they include human capital in the measure of wealth and allow for time-varying betas, finding a significant increase in the cross-sectional variation of returns that the modified CAPM is able to explain.

The above extension as well as many others have been explored by a number of scholars seeking to justify, challenge, or further improve the capital asset pricing model. Even with a strong theoretical justification and these extensions, however, the CAPM has nevertheless faced some difficulty in explaining real world asset prices and expected returns. The failure of the CAPM to adequately explain much of the variation in cross-sectional stock returns has motivated the discovery of other important risk factors, most notably size and value in Fama and French (1993). The Momentum factor was then described by Jegadeesh and Titman (1993) and the Fama-French-Carhart 4 Factor model introduced in Carhart (1997).

As the discussion over asset pricing models rages on, the mutual fund industry has appeared as a ripe area of empirical research regarding which asset pricing model investors actually rely on. The swathe of publicly available data on mutual funds, as well as their size in the world of investments - US Mutual Funds hold $17.7 trillion of assets as of the end of 2018 - has made the mutual fund performance data a solid bedrock for such empirical research.[1] Mutual funds also attract a spectrum of investors, including both retail and institutional clients. The popularity of mutual funds can be seen from the fact that in the US, around 45% of all households own funds, and $8.2 trillion of the entire $27.1 trillion retirement market is invested in mutual funds.[2]

The past two decades have garnered significant attention to mutual fund performance. The pioneering work of Berk and Green (2004) demonstrates, through a theoretical model with decreasing returns to scale with regards to fund size, that fund flows respond to manager performance, which is a signal of their skill, until in equilibrium competitive market forces drive the net assets of the fund up to the level where net alpha is zero. Although performance may not be persistent, Berk and Green (2004) argue that manager skill exists, is heterogeneous across managers, and can be identified by investors.

The assertion that mutual fund managers have skill that investors can identify, markets are competitive, and in equilibrium net alpha is zero led to a new way of empirically testing which asset pricing model investors most closely rely on. Berk and van Binsbergen (2016) show that in a competitive market with rational investors, mutual funds that generate alpha must either lead to higher fund fees or higher fund inflows, which would lead to decreasing returns to scale and eventually eradicate the positive alpha opportunity. Perhaps the most important implication of the paper is highlighting the ability of mutual fund data to help determine which asset pricing model investors are using.

As the focus has shifted to the primary drivers behind investor flows, the arsenal of potential asset pricing models and other performance metrics investors might be using to gauge fund manager skill has grown. Barber et al. (2016) consider not only the CAPM but also the

[1] 2019 Investment Company Fact Book: A Review of Trends and Activities in the Investment Company Industry

[2] Ibid.

Fama-French 3-Factor model, Fama-French-Carhart 4-Factor model, a 7-Factor model with the three industry factors of Pastor and Stambaugh (2002), and a 9-Factor model adding the profitability and investment factors of Fama and French (2015). Out of all the models considered, the CAPM appears to be most useful in explaining investor flows as there are greater flows to mutual funds with higher ranks based on CAPM alphas than to funds with higher ranks based on any of the other models considered. Barber et al. (2016) also find that when evaluating fund performance, investors appear to care the most about market risk and treat returns attributable to size, value, momentum, and industry factors as alpha.

Following the work of Berk and van Binsbergen (2016) and Barber et al. (2016) that demonstrated investors predominantly rely on the CAPM model when allocating capital to funds, others have come to challenge that result by demonstrating that Morningstar Ratings, in fact, have much greater explanatory power for investor flows. Del Guercio and Tkac (2008) are among the first to show that Morningstar rating changes, even absent changes in underlying fund performance, drive retail mutual fund flows. More recently, Evans and Sun (2020) have supported this finding by demonstrating that the Morningstar rating change of June 2002 affected aggregate risk adjustment by retail mutual fund investors. My paper results differ from this literature by considering not only mutual fund investors but also those in separate account composites, as well as looking at passively managed funds, considering the Sharpe ratio and Information ratio performance metrics, and breaking the datasets up by time period and investment style.

In their approach, Ben-David et al. (2020) rely on the data and methodology from Barber et al. (2016), with the important addition of also considering Morningstar data. They also perform an econometric test that addresses the critique first brought forward by Pastor et al. (2015), namely that regressions of flows on CAPM alphas will place the most weight on periods of high market volatility, during which flows may be driven by contagion or other market-level events rather than strictly fund manager skill assessments. Barber et al. (2016)'s takeaway is that because investors do not respond to changes in the market related component of return as much as to changes in other factor related components of return, they are using the CAPM model. Ben-David et al. (2020) dispute this finding by describing how Pastor et al. (2015)'s critique has an effect here. To correct for this, Ben-David et al. (2020) run a two-stage Fama MacBeth regression and find that the results do not show a lower beta on market-related component of return than on other factor-related components of return, that is investors respond to returns associated with market risk more closely to how they respond to value, size, and momentum factor returns. They conclude that investors do not use the CAPM model, or in fact any other factor model.

I also find that Morningstar ratings supersede factor models in explaining flows as in Ben-David et al. (2020), but extend this analysis by using a novel dataset, US Separate Account Composites from Morningstar. I find that the separate account composite investors also rely on Morningstar ratings more than on the CAPM or factor model alphas, but relatively less so than their Mutual Fund investor counterparts. This finding, corroborated on a novel dataset, is in line with prior results from Evans and Fahlenbrach (2012), who look at retail-institutional mutual fund pairs and find that institutional investors are more sensitive to high fees and poor risk-adjusted performance, and in line with Evans and Sun (2020) who show that flows to institutional separate accounts strongly correlate with the Fama French 3 Factor alphas but not with the CAPM alphas.

I also extend the literature by looking at other easily accessible and widely understood popular measures of investment fund performance, the Sharpe ratio and Information ratio. I include these two measures in my tests because of how easily they can be understood by investors and how commonly they are included in fund performance summary or marketing materials. There is also prior research backing for similarly easily calculated and easily understood metrics, such as active share described in Cremers and Petajisto (2009). I find that the Information ratio performs better than the Sharpe ratio in predicting flows, but both demonstrate less predictive power than the CAPM or Morningstar ratings. I then propose some possible explanations and mechanisms for this and pave the way for future theoretical research in this direction.

Finally, my findings regarding the separate account composites performance extend the empirical work of Gerakos et al. (2019), who use a $17 trillion dataset of institutional accounts to document outperformance over the 2000-2012 period. For my separate account composite sample, I also document relative outperformance over the first period (1991-2011) but find underperformance over the second period (2012-2020), indicating that there is no conclusive evidence for consistent separate account outperformance.

3 Data

3.1 US Separate Accounts (Composite) Data

The main data source that I rely on in my analysis is the US Separate Accounts database from Morningstar. This is a different and importantly unique data source from others commonly used in the mutual fund literature since because it includes only separately managed accounts of large and sophisticated investors. On average, these accounts have a minimum investment of $10 million. Furthermore, investing in a separate account allows investors to customize their strategy with exclusion of Tobacco stocks or tighter individual stock position limits for instance. The separate account investors are also protected from the risk of inflows and outflows by other investors, as occurs in a pooled investment vehicle such as a mutual fund. These factors combine to paint a picture of separate account investors as some of the largest, most sophisticated, and well respected in the entire US investing landsape.

The separate accounts database includes monthly gross and net returns, net assets, Morningstar ratings, and fund characteristics for of separate account composites. A composite is defined as the aggregation of one or more fully discretionary portfolios managed according to a similar investment mandate, and is the primary vehicle that firms use in presenting performance to prospective clients. All composite performance data is typically compliant with strict performance reporting requirements such as the Global Investment Performance Standards (GIPS) to guard investors against misleading information and facilitate consistent information and comparisons between various composites.

In my dataset, I include all composites that have a base currency of USD, whose Global Broad Category Group is Equity, whose US Category Group is US Equity, and whose Primary Prospectus Benchmark is managed by on of the top US Equity benchmark providers of S&P 500, Russell, or Nasdaq. Composites are formed from their constituents, so that the composite returns are the weighted average returns of the separate accounts included in a given composite. Composite assets are calculated as the sum of the assets of the composite constituents.

My dataset includes a total of 3,787 composites, each with on average 188 months of observations for a total of 711,221 fund-month observations. Of those, the vast majority are classified as Active funds, with a minority classified as Enhanced Index or Passive Funds. The Active funds include 3,271 composites, each with approximately 10 years of monthly data for a total of 637,334 fund-month observations. Enhanced Index funds constitute a total of 17,328 fund-month observations while Passive funds include 18,707 fund-month observations. The remaining 37,852 fund-month observations of the full dataset are not marked as either Active, Enhanced Index, or Passive funds.

The variables included in the dataset are monthly gross returns, monthly net returns, monthly net assets, and monthly Morningstar ratings as well as fund characteristics such as Morningstar Category (e.g. Mid-Cap Value or Small-Cap Growth) and Management Approach (Active, Enhanced Index, or Passive). I also have indicators relating to the institutional or retail nature of the separate account composites. The Product Focus indicator, for instance, refers to whether the product's focus is institutional or retail (where institutional is defined by Morningstar as having a minimum investment of $100,000), while the Minimum Investment indicator states the minimum investment amount for a strategy account within the given composite. The average Minimum Investment across the composites is around $10 million. This information is useful for my latter analysis of how retail investors behave differently from institutional investors, described in more detail in Section 5. There are several additional characteristics of interest, such as whether a fund is Open for Investment and its Branding Name, which is equivalent to the overarching firm that operates all of the constituent funds within each of its composites.

In Table 1, I present the descriptive statistics for the US Separate Accounts Composites - Active dataset.

Table 1: Descriptive Statistics. US Composite Accounts, Active

Descriptive statistics of the observation data sample for US composite accounts, sorted into buckets by Morningstar Rating. This table summarizes the main descriptive statistics for the US Separate Account Composites sample of active funds over the time period 01/01/1991 - 09/30/2020. Observations refers to fund-month observations, and all statistics are computed as averages over these observations.

	Morningstar Rating						
	1 Star	2 Stars	3 Stars	4 Stars	5 Stars	Rating NA	All
Fund-month observations	14,152	53,715	99,069	68,793	21,836	379,769	637,334
Fund size ($million)	626.66	849.57	1,402.74	1,968.13	2,570.27	1,561.62	1,534.75
Fund age (years)	14.73	15.04	14.84	14.24	12.54	6.99	10.03
Fund flow	-1.64%	-1.12%	0.13%	1.23%	3.53%	1.91%	0.80%
Mkt-Adj Return	-0.26%	-0.14%	-0.06%	0.00%	0.07%	0.24%	0.12%
Excess Return	0.60%	0.68%	0.76%	0.83%	0.90%	0.77%	0.77%
Ret. Volatility (1yr)	5.12%	4.60%	4.44%	4.30%	4.36%	4.59%	4.54%
Ret. Volatility (5yr)	5.40%	4.84%	4.62%	4.45%	4.34%	4.96%	4.78%
Market beta	1.03	1	0.98	0.96	0.94	0.97	0.97
Size beta	0.34	0.26	0.24	0.24	0.23	0.21	0.23
Value beta	0.09	0.09	0.1	0.1	0.08	0.09	0.09
Momentum beta	0.01	0.02	0.02	0.02	0.02	0.01	0.01
Fraction of positive flows	27.83%	28.96%	33.56%	43.76%	60.60%	49.75%	41.36%

As can be seen from Table 1, average beta loading for the funds is very close to 1, the average size loading is around 0.2 and the average value and momentum loadings are relatively low at 0.08 and 0.02 respectively. It is worth noting that the US Separate Account Composites dataset includes a large number of US Small Cap universe funds, so it is to be expected that the sample overall would have a sizable size loading. This is due to the fact that funds benchmarked to Russell 2000 or other small cap indices will by construction hold on average smaller stocks than the general US market.

Although most of the composites do not have Morningstar ratings, I still have a sizable sample of those that do and can glean interesting observations from the above table. For instance, separate account composites with higher Morningstar ratings tend to be larger in fund size, with higher market-adjusted and excess of risk-free rate returns and relatively lower 1 year and 5 year standard deviations of return. This makes sense since by construction, the Morningstar rating takes fund returns and volatility into account; more details on the Morningstar rating and the aspects of a fund that it considers can be found in the Data Points section.

Interestingly, funds with higher Morningstar ratings do not have higher average age. They tend to have lower market and size betas, while also a slightly lower value beta and a slightly higher momentum beta.

Perhaps most interestingly, the average fund flows and fraction of positive flows correlate very strongly with the Morningstar ratings. Average flows for 1-star and 2-star funds are negative, while average flows for 3-star, 4-star, and 5-star funds are positive and increasing with the number of stars, reaching a high of 9.03% for the highest-ranked 5-star funds. The fraction of positive flows for each star category shows a similar pattern: it is monotonically increasing in the number of stars, with a range from 31.15% for 1 star funds to 65.32% for 5 star funds. This can be construed as some supporting evidence towards the claim that even investors in separate accounts care about Morningstar ratings and direct their flows accordingly. However, causality in the table is not explicit, and there are many other variables, such as CAPM alphas or other measures of risk-adjusted performance, that may simply be correlated with Morningstar ratings but be the true driving force behind investors flows. I examine tests for this in Section 5.

3.2 US Mutual Funds Data

The second dataset I use is the US Mutual Fund dataset from Morningstar. The full sample period considered is 01/01/1980 - 09/30/2020. I further split up this data sample by Style (Value, Growth, and Blend), Management Approach (Acitve or Passive), and Time Period (1991-2011 and 2012-2019) for certain portions of the analysis as will be described in more detail in Section 5. For instance when testing the hypotheses brought forward by Barber et al. (2016), I consider only the time period January 1991 - December 2011 to be consistent with their analysis. I consider both currently existing and past funds in my analysis to avoid survivorship bias.

For each US Mutual Fund in my data, I gather the following data points by share class: monthly gross returns, monthly net returns, monthly net assets, month net expense ratios, monthly Morningstar ratings, annual turnover, and share class characteristics such as Fund Name, Manager Name, Index Fund flag, Oldest Share Class flag, and Morningstar Category.

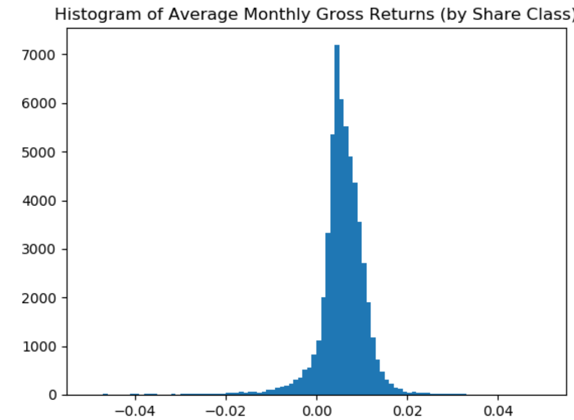
Monthly net returns in Morningstar, referred to as monthly total returns, are calculated by taking the change in monthly net asset value, reinvesting all income and capital-gains distributions during that month, and dividing by the starting NAV. Reinvestments are made using the actual reinvestment NAV, and daily payoffs are reinvested monthly. Net returns do not adjust for sales charges such as front-end loads, deferred loads and redemption fees but do account for management fees, administrative fees, and 12b-1 fees.

Gross monthly return is calculated by Morningstar by adjusting the monthly net return for the share class by the share class level fees prevailing at that time, which come from the most recent net expense ratio. This return measure is thus gross of any expenses paid. For periods where Morningstar does not have the prevailing fees for the share class, gross returns are not calculated. To have a visual representation of the data, Figure 1 depicts a histogram of average monthly gross returns (averaged over time by share class) and a time series of cumulative average monthly gross returns (where average monthly gross return is averaged over share classes for any given month).

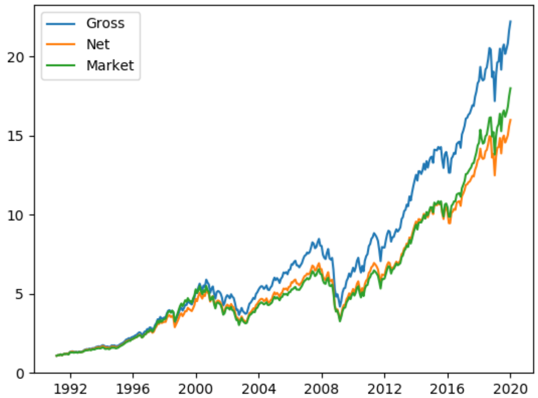
Monthly net assets for each share class are defined as the monthly share-class level of total net assets. Fund net assets are the sum of share class net assets for all share classes in the fund.

The Net Expense Ratio is the percentage of fund assets used to pay for operating expenses and management fees, including 12b-1 fees, administrative fees, and all other asset-based costs incurred by the fund, except brokerage costs. Fund expenses are reflected in the fund's NAV. Sales charges are not included in the expense ratio.

The turnover ratio measures a fund's trading activity. It is computed by taking the minimum of the purchases or sales of a share class and dividing it by the average monthly net assets over



(a) Histogram of Average Gross Returns



(b) Time Series of Cumulative Gross, Net, and Market Returns

Figure 1: Figure (a) shows a histogram of average monthly gross returns (by share class) for the actively managed US Mutual Fund dataset. Average monthly gross returns are computed for each share class separately, averaging over the lifetime of that share class. Figure (b) shows a time series of cumulative gross returns, net returns, and market returns for the actively managed US Mutual Fund dataset over 1991-2019. The gross and net return time series are computed by first taking the average gross and net monthly return across all share classes in each month, and then cumulating these average returns over time.

the period. Securities with maturity less than a year are excluded from the annual turnover measure. Morningstar does not calculate turnover ratios, but rather gathers them from the financial highlights of funds' annual reports. The turnover ratio thus loosely corresponds to the percent of a share class's holdings that have changed over a year. Low turnover would typically correspond to a buy and hold strategy while high turnover would indicate a more active strategy.

The Morningstar Rating brings returns, risk, and load adjustments together into a single rating. Morningstar ratings are computed over three general time periods: 3 years, 5 years, and 10 years. For a given time period, to determine a fund's Morningstar rating its risk-adjusted return is plotted on a bell curve within its category group. The funds that score in the top 10% of the category are classified as 5 star funds; those in the next 22.5% receive 4 stars; the middle 35% are classified as 3 stars; the next 22.5% receive 2 stars; and the bottom 10% receive 1 star. The Overall Morningstar Rating, which is the Morningstar rating that I use in my analysis, is a weighted average of the 3 star, 5 star, and 10 star ratings. For funds with between 3 years and 5 years of return history, the Overall Morningstar rating is simply the 3 year rating. For funds with between 5 years and 10 years or return history, the Overall Morningstar rating is calculated as the sum of 60% of the 5 year rating and 40% of the 3 year rating. Finally for funds with over 10 years of return history, the Overall Morningstar rating is calculated as the sum of 50% of the 10 year rating, 30% of the 5 year rating and 20% of the 3 year rating. More details regarding the Morningstar rating are in the following Section 4.

After gathering all of the above data points, I clean the data which is described in more detail in Appendix B: Database Cleanup and Merge. I follow prior literature methodology in aggregating up to fund level and removing data points with no return data or flows below 90% or above 1000%.

In Table 2, I present the descriptive statistics for the actively managed US Mutual Funds dataset.

Table 2: Descriptive Statistics. US Active Mutual Funds

Descriptive statistics of the observation data sample for US active mutual funds, sorted into buckets by Morningstar Rating. This table summarizes the main descriptive statistics for the US Mutual Funds sample of active funds over the time period January 1991 - December 2019. Observations refers to fund-month observations, and all statistics are computed as averages over these observations.

	Morningstar Rating						
	1 Star	2 Stars	3 Stars	4 Stars	5 Stars	Rating NA	All
Fund-month observations	23,243	85,901	150,463	108,629	39,900	55,801	463,937
Fund size ($million)	489.04	815.39	1,555.23	2,470.44	3,323.59	284.21	1578.33
Fund age (years)	9.96	10.53	10.65	10.17	8.51	1.33	9.18
Fund flow	-1.63%	-1.02%	-0.28%	0.96%	3.14%	4.94%	0.73%
Mkt-Adj Return	-0.40%	-0.13%	0.01%	0.18%	0.52%	0.11%	0.06%
Excess Return	0.22%	0.54%	0.70%	0.85%	1.06%	0.67%	0.71%
Ret. Volatility (1yr)	5.33%	4.62%	4.33%	4.28%	4.61%	4.57%	4.47%
Ret. Volatility (5yr)	5.72%	5.04%	4.73%	4.57%	4.55%	5.42%	4.80%
Market beta	1.05	1.01	0.99	0.97	0.97	1.00	0.99
Size beta	0.31	0.23	0.19	0.18	0.21	0.22	0.21
Value beta	-0.01	0.03	0.05	0.07	0.05	0.02	0.04
Momentum beta	0.00	0.01	0.01	0.01	0.01	0.01	0.01
Fraction of positive flows	18.78%	22.42%	32.77%	51.21%	71.97%	69.62%	42.27%

Table 2 includes many of the same patterns observed in Table 1 for US Separate Account

Composites, but also some important differences. I first focus on the similarities.

The average beta loading for the funds is again very close to 1, the average size loading is around 0.2 and the average value and momentum loadings are relatively low. These loadings are very comparable to those seen for the separate account composites, with the notable exception of the value loading being slightly higher for separate account composites than for mutual funds.

The pattern of monotonically higher flows and higher fraction of positive flows for funds with higher Morningstar ratings remains similarly intact. Average flows for 1-star, 2-star, and 3-star funds are negative, while average flows for 4-star and 5-star funds are positive and increasing with the number of stars, reaching a high of 3.14% for the highest-ranked 5-star funds. The fraction of positive flows for each star category shows a similar pattern: it is monotonically increasing in the number of stars, with a range from 18.78% for 1 star funds to 71.97% for 5 star funds. This can be viewed as supporting evidence that mutual fund investors care about Morningstar ratings and direct their flows accordingly. However, as discussed before causality here is not implicit. Although the general pattern of increasing fraction positive flows to funds with higher Morningstar ratings is similar between US Mutual Funds and US Separate Account Composites, I also note one important difference: the range between 1 star funds' and 5 star funds' fractions of positive flows is much larger for US Mutual Funds, indicating that mutual fund investors potentially react to Morningstar ratings more than their separate account counterparts.

As for separate account composites, mutual funds with higher Morningstar ratings also tend to be larger in fund size, with higher market-adjusted and excess of risk-free rate returns and relatively lower 1 year and 5 year standard deviations of return. However compared to separate account composites, mutual funds exhibit simultaneously lower returns and higher volatility, giving some credence to the commonly postulated belief that institutional investors are better at identifying lower fee and higher risk-adjusted return opportunities, or that their flows guide fund managers more towards that direction. This finding corroborates Evans and Fahlenbrach (2012), who find that institutional investors' sensitivity to high fees and poor risk-adjusted performance is higher than for their retail twins.

Finally, from Table 2 I see that as for separate account composites, mutual funds with higher Morningstar ratings do not have higher average age. They also tend to have lower market and size betas.

3.3 Computing Flows and Model Alphas

Armed with the full US Mutual Fund dataset, I next turn to combining the share classes up to the fund level. Many mutual funds offer multiple share classes, which represent claims on the same underlying assets but have different fee structures. In Morningstar, share classes of the same fund are represented by the same FundId. Consequently I aggregate the share classes of the same fund by FundId, so that my final sample contains fund-level returns, assets, and other characteristics. More specifically, I compute a fund's net assets by summing net assets across the fund's share classes, and I compute a fund's gross and net returns, expense ratios, turnover, and Morningstar ratings by asset-weighting across share classes.

Each month to calculate fund level total net assets (referred to as TNA for brevity), I sum the total net assets for each share class of the fund for that month. To calculate the fund level gross monthly return, I asset-weight the share class gross monthly returns and proceed analogously for net monthly returns, expense ratios, turnover, and Morningstar ratings. For a fund F at time t I then have:

$$\begin{aligned}
\mathrm{TNA}_t^F &= \sum_{\text{Fund F's Share Classes i}} \mathrm{TNA}_t^i \\
R_t^{F,Gross} &= \sum_{\text{Fund's Share Classes i}} R_t^{i,Gross} * \frac{\mathrm{TNA}_t^i}{\mathrm{TNA}_t^F} \\
R_t^{F,Net} &= \sum_{\text{Fund's Share Classes i}} R_t^{i,Net} * \frac{\mathrm{TNA}_t^i}{\mathrm{TNA}_t^F} \\
\text{Expense Ratio}_t^F &= \sum_{\text{Fund's Share Classes i}} \text{Expense Ratio}_t^i * \frac{\mathrm{TNA}_t^i}{\mathrm{TNA}_t^F}
\end{aligned}$$

$$\text{Turnover}_t^F = \sum_{\text{Fund's Share Classes i}} \text{Turnover}_t^i * \frac{\text{TNA}_t^i}{\text{TNA}_t^F}$$

$$\text{Morningstar Rating}_t^F = \sum_{\text{Fund's Share Classes i}} \text{Morningstar Rating}_t^i * \frac{\text{TNA}_t^i}{\text{TNA}_t^F}$$

where TNA_t^F are fund F's total net assets at time t, TNA_t^i are share class i's total net assets at time t, $R_t^{F,Gross}$ is fund F's gross monthly return at time t, $R_t^{i,Gross}$ is share class i's gross monthly return at time t, $R_t^{F,Net}$ is fund F's net monthly return at time t, and $R_t^{i,Net}$ is share class i's net monthly return at time t.

Having aggregated up the data points to fund level, I then leave only one observation per fund. I elect to leave only the oldest share class at any point in time of a fund.

From the fund level assets and net returns, in line with prior literature I obtain flows to fund F at time t according to the following formula:

$$\text{Flows}_t^F = \frac{TNA_t^F}{TNA_{t-1}^F} - (1 + R_t^{F,Net})$$

where Flows_t^F are the calculated flows to fund F at time t, TNA_t^F are fund F's total net assets at time t, TNA_{t-1}^F are fund F's total net assets at time t-1, and $R_t^{F,Net}$ is fund F's net monthly return between month t-1 and month t.

For the next step of computing a fund's factor loadings over time, I use the fund's fund-level monthly net returns computed as explained above, the fund's benchmark returns data gathered from Morningstar, and value, size, and momentum factor returns, market return, and risk-free rate return data from Ken French's data library.[3]

I compute the $\beta_t^{F,CAPM}$ time t loading of fund F on the CAPM market beta by running a rolling 60 month regression of the fund's monthly net returns on monthly factor returns using data from months $\tau = t-1, t-60$ as follows:

[3]The market return, risk-free rate return, and value, size, and momentum factor returns can be found on Ken French's data library website at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

$$(R_\tau^{F,Net} - R_\tau^{RF}) = \alpha_t^{F,CAPM} + \beta_t^{F,CAPM}(R_\tau^{Mkt} - R_\tau^{RF}) + \epsilon_t^F \tag{1}$$

where $R_\tau^{F,Net}$ is fund F's net monthly return in month τ, R_τ^{RF} is the risk-free rate return in month τ, and R_τ^{Mkt} is the market return in month τ.

Similarly, to obtain fund F's time t loadings on market, size, and value factors in the Fama-French 3 Factor model (call these loadings $\beta_t^{F,3F}$, $s_t^{F,3F}$, and $h_t^{F,3F}$ respectively), I run the following rolling window regression at each time t using fund F's net returns and the Fama-French 3 Factor returns data for the months $\tau = t-1, t-60$:

$$(R_\tau^{F,Net} - R_\tau^{RF}) = \alpha_t^{F,3F} + \beta_t^{F,3F}(R_\tau^{Mkt} - R_\tau^{RF}) + s_t^{F,3F}(SMB_\tau^{3F}) + h_t^{F,3F}(HML_\tau^{3F}) + \epsilon_t^F \tag{2}$$

where $R_\tau^{F,Net}$ is fund F's net monthly return in month τ, R_τ^{RF} is the risk-free rate return in month τ, R_τ^{Mkt} is the market return in month τ, SMB_τ^{3F} is the size factor return for the 3 Factor model in month τ, and HML_τ^{3F} is the value factor return for the 3 Factor model in month τ.

For the 4 Factor model, the market, size, value, and momentum factor loadings ($\beta_t^{F,4F}$, $s_t^{F,4F}$, $h_t^{F,4F}$, and $u_t^{F,4F}$ respectively) for fund F at time t are estimated analogously from:

$$\begin{split}(R_\tau^{F,Net} - R_\tau^{RF}) = \alpha_t^{F,4F} + \beta_t^{F,4F}(R_\tau^{Mkt} - R_\tau^{RF}) + s_t^{F,4F}(SMB_\tau^{4F}) \\ + h_t^{F,4F}(HML_\tau^{4F}) + u_t^{F,4F}(UMD_\tau^{4F}) + \epsilon_t^F\end{split} \tag{3}$$

where $R_\tau^{F,Net}$ is fund F's net monthly return in month τ, R_τ^{RF} is the risk-free rate return in month τ, R_τ^{Mkt} is the market return in month τ, SMB_τ^{4F} is the size factor return for the 4 Factor model in month τ, HML_τ^{4F} is the value factor return for the 4 Factor model in month τ, and UMD_τ^{4F} is the momentum factor return for the 4 Factor model in month τ.

Net alphas for fund F at time t are then computed as the fund's monthly net return less the product of that month's estimated factor loadings (using data for months t-60 to t-1 as explained above) and the current month's factor returns. For the CAPM, 3 Factor, and 4 Factor models the alpha computations for fund F at time t are as follows:

$$\hat{\alpha}_t^{F,CAPM} = (R_t^{F,Net} - R_t^{RF}) - \hat{\beta}_t^{F,CAPM}(R_t^{Mkt} - R_t^{RF}) \tag{4}$$

$$\hat{\alpha}_t^{F,3F} = (R_t^{F,Net} - R_t^{RF}) - \left[\hat{\beta}_t^{F,3F}(R_t^{Mkt} - R_t^{RF}) + \hat{s}_t^{F,3F}(SMB_t^{3F}) + \hat{h}_t^{F,3F}(HML_t^{3F})\right] \tag{5}$$

$$\begin{aligned}\hat{\alpha}_t^{F,4F} = (R_t^{F,Net} - R_t^{RF}) - \Big[&\hat{\beta}_t^{F,4F}(R_t^{Mkt} - R_t^{RF}) + \hat{s}_t^{F,4F}(SMB_t^{4F}) \\ &+ \hat{h}_t^{F,4F}(HML_t^{4F}) + \hat{u}_t^{F,4F}(UMD_t^{4F})\Big]\end{aligned} \tag{6}$$

where $R_t^{F,Net}$ is fund F's net monthly return in month t, R_t^{RF} is the risk-free rate return in month t, R_t^{Mkt} is the market return in month t, SMB_t^M is the size factor return for factor model M in month t, HML_t^M is the value factor return for factor model M in month t, UMD_t^M is the momentum factor return for factor model M in month t, and $\beta_t^{F,M}$, $s_t^{F,M}$, $h_t^{F,M}$, and $u_t^{F,M}$ are the estimated factor model M loadings for fund F at time t.

Finally, I compute weighted alphas from estimated monthly alphas according to the methodology described in Barber et al. (2016), relying on the prior 18 months of alpha data to compute the weighted alpha. The weighted alpha for fund F at time t using model M is computed as follows:

$$\alpha_t^{F,M,Weighted} = \frac{\sum\limits_{s=1}^{18} e^{-\lambda(s-1)}\hat{\alpha}_{t-s}^{F,M}}{\sum\limits_{s=1}^{18} e^{-\lambda(s-1)}} \tag{7}$$

where $\hat{\alpha}_{t-s}^{F,M}$ is the estimated alpha for fund F at time t-s using model M, and λ is a decay parameter in the return-flow relation over time. Barber et al. (2016) calibrate this decay parameter to $\lambda = 0.20551497$; this decay parameter value is later used by Ben-David et al.

(2020) and I similarly use it in my calculation of weighted alphas.

4 Other Common Performance Metrics

4.1 Morningstar Rating

The Morningstar Rating is an important consideration for fund flows because it is a very widely-used measure by investors and advisors alike. Investors can use it as a quick gauge of fund quality as it combines fund returns, risk, and ranking against other funds within a category into a single easy-to-interpret metric. For advisors, the Morningstar rating can act as a salient selling point of a fund to investors. Consultants can use it to pitch funds that will be included on retirement investment option platforms.

Since the Morningstar rating compares funds against other funds in a particular Morgningstar category, it can also be especially useful in choosing a small number of funds from many available in a certain asset class or subclass. It was introduced in 1985, and has remained an integral measure for fund ratings and marketing material. As I will show, it also has power in explaining investor flows and consequently appears to be taken seriously by investors.

While the methodology of the Morningstar rating has changed somewhat over the years, the main aspects of it have remained the same. In essence, it compares funds within broad category groups against each other based on risk-adjusted return metrics. The categories were initially more broad, such as equity or fixed income, and have since become more granular to accommodate growing demand for funds fulfilling more specific portfolio allocation needs. As investors began to appreciate the correlations between their various investments more, funds were increasingly chosen as portfolio components rather than stand-alone investments and the need for more granular Morningstar ratings became apparent. Evans and Sun (2020) document how this rating methodology change affected the aggregate risk adjustment made by retail investors.

The formula for the Morningstar ratings calculated over 3 year, 5 year, and 10 year time periods is:

$$MRAR(\gamma) = \left[\frac{1}{T}\sum T(1+ER_t)^{-\gamma}\right]^{\frac{12}{\gamma}} - 1 \tag{8}$$

where ER_t is the geometric excess net return in month t, T is the number of months in the time period, and γ is a risk tolerance parameter that Morningstar analysts have calibrated to $\gamma = 2$ to be consistent with their measurements of risk tolerances of typical retail investors. As mentioned in Section 3, the Overall Morningstar rating which I use in my analysis is a weighted combination of the 3 year, 5 year, and 10 year Morningstar ratings calculated according to the formula above.

Table 3: Morningstar Ratings: Data Availability for US Mutual Funds

Number of fund-month observations by dataset. Datasets are all sub-samples of US Mutual Funds. Where time period is unspecified, the full time period 1991 - 2019 is assumed. Where time period is specified, 1st Period refers to 1991 - 2011 and 2nd Period refers to 2012 - 2019.

Morningstar Rating Observations							
	1 Star	2 Stars	3 Stars	4 Stars	5 Stars	Rating NA	All
Active Funds	23,243	85,901	150,463	108,629	39,900	55,801	463,937
Active Funds - 1st Period	14,681	55,287	93,083	69,964	27,657	44,893	305,565
Active Funds - 2nd Period	8,562	30,614	57,380	38,665	12,243	10,908	158,372
Value Style Active Funds	6,224	22,762	41,442	29,879	10,341	14,264	124,912
Growth Style Active Funds	10,443	37,931	67,482	48,210	18,727	22,757	205,550
Blend Style Active Funds	5,446	22,284	37,610	27,948	10,011	15,208	118,507
Passive Funds	203	2,293	12,143	8,799	1,813	4,553	29,804

As mentioned, my data sample covers the 1991-2019 period and I break it up into three further time periods in my analysis: the entire time period, 1991-2011 to be consistent with prior work, and 2012-2019 to see how investor revealed preferences for asset models have changed in the most recent decade. Over all three data periods, as will be discussed in more detail in Section 5, the Morningstar rating appears to be the most significant driver of investor flows though the 4 Factor alpha becomes a relatively more important driver of flows than before. I also look at the difference between funds within a certain investment style, such as Value, Growth, and Blend. Finally, I consider the sub-sample of passively managed mutual funds. An overview of the number of fund-month observations I have for these various datasets is shown in Table 3.

By breaking up my data sample according to these sub-samples, I am able to see how much the Morningstar ratings drive flows for funds with different management approaches (active vs passive) and different style (value vs growth vs blend), and how that has changed over time. All of these tests and results are summarized in Section 5.

4.2 Sharpe ratio

If the Morningstar Rating, a measure of risk-adjusted performance, is so important to investors, would the Sharpe ratio also be a significant determinant of flows? In this paper, I seek to find that out.

The main reasons for why the Sharpe ratio is an important and potentially significant metric for investors when allocating capital to funds are that the Sharpe Ratio is a very well-known, easily understood, and generally accepted measure of a fund's performance. Since it was first introduced by William Sharpe in Sharpe (1966), it has gained widespread popularity and acceptance by the practitioner and academic communities alike.

While the Sharpe ratio and the Morningstar rating are both risk-adjusted measures of performance, the Morningstar rating differs from the Sharpe ratio because it gives more weight to downside variation and does not make any assumptions about the distribution of excess returns. The Sharpe ratio, by contrast, uses standard deviation which is a symmetric variation measure.

My data sample contains monthly Morningstar rating observations for the period 1991 - 2019, and consequently I calculate monthly Sharpe ratio data over the same period. As is customary, I do not calculate Sharpe ratios for funds with fewer than 12 consecutive months of observations. For funds with at least 12 months of observation, the Sharpe ratio for fund F at time t is calculated using all available fund returns data up until time t according to the following formula:

$$SR_t^F = \frac{\bar{R}_t^F - \bar{R}_t^{RF}}{\sigma_t^{F,ann}} \tag{9}$$

where t is the fund's age in months, SR_t^F is fund F's annualized Sharpe Ratio using return observations from month 1 to month t of the fund's history, $R_t^{F,Net}$ is fund F's net monthly return in month t, $\bar{R}_t^F = \left[\prod_{s=1}^t (1 + R_s^{F,Net})\right]^{12/t}$ is the annualized geometrically compounded net return of fund F using return data from months 1 to t of the fund's history, $\bar{R}_t^{RF} = \left[\prod_{s=1}^t (1 + R_s^{RF})\right]^{12/t}$ is the annualized geometrically compounded return of the risk-free rate[4] using return data from months 1 to t of the fund's history, and $\sigma_t^{F,ann}$ is the annualized standard deviation of fund F's net monthly returns computed using returns from months 1 to t of the fund's history.

4.3 Information ratio

The rationale for using the Information ratio as a performance metric that could help explain investor flows is much the same as that behind using the Sharpe ratio. The Information ratio is a very well known, straightforward, well-understood and well-regarded measure of fund performance. While the Sharpe ratio compares a fund to the risk-free rate benchmark, the Information ratio measures fund performance against its designated benchmark which is likely to be more relevant for the fund's investors. I compute the Information ratio as follows:

$$IR_t^F = \frac{\bar{R}_t^F - \bar{R}_t^{Bmk}}{TE_t^{F,ann}} \tag{10}$$

where t is the fund's age in months, IR_t^F is fund F's annualized Information Ratio using return observations from month 1 to month t of the fund's history, $R_t^{F,Net}$ is fund F's net monthly return in month t, $\bar{R}_t^F = \left[\prod_{s=1}^t (1 + R_{s,Net}^F)\right]^{12/t}$ is the annualized geometrically compounded net return of fund F using return data from months 1 to t of the fund's history, $\bar{R}_t^{Bmk} = \left[\prod_{s=1}^t (1 + R_s^{Bmk})\right]^{12/t}$ is the annualized geometrically compounded return of the fund's

[4] The risk-free rate is taken to be the 1-Month Treasury Bill rate from Ibbotson Associates, as sourced from the Ken French data library website at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

benchmark using return data from months 1 to t of the fund's history, and $TE_t^{F,ann}$ is the annualized tracking error, defined as the annualized standard deviation of the difference between fund F's net monthly returns and fund F's benchmark returns computed using returns from months 1 to t of the fund's history. Note that for more exact measures of tracking error using daily returns is preferred, so using monthly returns in this case serves as an approximation.

5 Results

5.1 Flows Sign Test for US Mutual Funds

I first perform the flows sign asset pricing model test initially introduced and justified in Berk and van Binsbergen (2016). Rather than using prices and returns, Berk and van Binsbergen (2016) propose to test which asset pricing model investors use by looking at the quantities of inflows and outflows to different mutual funds. By this logic, the model that best explains flows would be the model that investors are most likely using.

The core idea is that mutual fund investors compete with each other to allocate capital into positive net present value opportunities. When adjusting for risk using the correct asset pricing model, funds with positive alphas are then exactly those with positive net present value opportunities. Consequently these funds should receive positive fund flows. Therefore by considering how well the signs of alphas match the directions of flows, it should be possible to deduce which asset pricing model investors are using. To test the asset pricing models using flows, Berk and van Binsbergen (2016) develop a simple test that uses flow sign and model alpha sign to infer the model closest to the one investors actually use in their capital allocation decisions.

The test works as follows. First, let F_t^i be fund i's flow in time t and $A_t^{i,M}$ be fund i's weighted model M alpha at time t. Note that since the alpha is weighted, it relies on data points of the previous 18 months and consequently only depends on past data, i.e. it is known to investors at time t. Now, define the sign function formally as follows:

$$\text{sign}(x) = \begin{cases} \frac{x}{|x|} \text{ if } x \neq 0 \\ 0 \text{ if } x = 0 \end{cases}$$

Next, consider the following regression:

$$\text{sign}(F_t^i) = \beta_0^M + \beta_1^M \text{sign}(A_t^{i,M}) + \epsilon_t^i \tag{11}$$

where $\text{sign}(F_t^i)$, the dependent variable, is the sign of Flows to fund i in month t and $\text{sign}(A_t^{i,M})$ is the sign of the model M weighted alpha of fund i in month t. If investors use an asset pricing model M, then the sign of the flow and the sign of the weighted model alpha should be the same.

In Lemma 2 of their paper, Berk and van Binsbergen (2016) further show that the estimated regression coefficient $\hat{\beta}_1^M$ can be expressed as:

$$\hat{\beta}_1^M = \frac{cov(\text{sign}(F_t^i), \text{sign}(A_t^{i,M})}{var(\text{sign}(A_t^{i,M}))} =$$

$$= P\Big[\text{sign}(F_t^i) = 1|\text{sign}(A_t^{i,M}) = 1\Big] + P[\text{sign}(F_t^i) = -1|\text{sign}(A_t^{i,M}) = -1] - 1 =$$

$$= P\Big[\text{sign}(F_t^i) = 1|\text{sign}(A_t^{i,M}) = 1\Big] - P\Big[\text{sign}(F_t^i) = 1|\text{sign}(A_t^{i,M}) = -1\Big]$$

Consequently, the frequency with which the flow sign and the model alpha sign agree with each other can be calculated from the estimated $\hat{\beta}_1^M$ coefficient as follows:

$$\frac{\hat{\beta}_1^M + 1}{2} = \frac{P\left[\text{sign}(F_t^i = 1)|\text{sign}(A_t^{iM}) = 1\right]}{2} + \frac{P\left[\text{sign}(F_t^i) = -1|\text{sign}(A_t^{iM}) = -1\right]}{2} \tag{12}$$

It is thus possible to run an asset pricing model test by looking at the model alpha sign coefficient of the regression in Equation 11, and using it to calculate the percent of times that the sign of flows is predicted by the signs of alphas of various models such as the CAPM, Fama-French 3 Factor model and Fama-French-Carhart 4 Factor model. The empirical results of Berk and van Binsbergen (2016) find that CAPM alphas match flows better than other factor models and better than market-adjusted, benchmark-adjusted, and excess of risk-free rate returns. They thus conclude that the CAPM is the closest to the true asset pricing model that investors use.

In a follow-up analysis, Ben-David et al. (2020) add a simple heuristic measure as an indicator variable that a fund's Morningstar rating is greater than or equal to a certain number of stars. For a fund F at time t, the 3 star Morningstar heuristic is an indicator variable equal to 1 if fund F's time t Morningstar rating is greater than or equal to 3 stars and equal to -1 otherwise. The 4 star and 5 star Morningstar indicator variables work analogously. Ben-David et al. (2020) test whether these Morningstar rating indicator variables predicts the direction of flows better than the factor model alpha signs do, and find confirmation it is so.

In my research, I add the Sharpe ratio and Information ratio to the possible predictors of flows that I test. Both are frequently presented to investors in fund prospectuses, and are therefore potential candidates for metrics that drive flows. My results of this comprehensive flow sign test are presented in Table 4. Following the approach of Berk and van Binsbergen (2016), I double cluster standard errors by fund and time.

I find, as shown in the univariate tests in Table 4, that the Morningstar rating 5 star indicator predicts the sign of flows far more often than the CAPM alpha or other model alphas do. While the predictability of the Morningstar indicator variables decreases as the number of stars that the indicator variables are based on decreases, even the Morningstar 3 star indicator predicts flows nearly 3% more often than the CAPM alpha.

Furthermore, in a model pairwise horserace, the difference between the Morningstar indicator variables and the model alphas in predicting flow sign is very significant. In fact, Morningstar rating 5 star, 4 star, and 3 star indicators all perform better than all other performance metrics considered in the test at the 1% significance level.

Table 4: Flow Sign Test - US Mutual Funds, Active, Full Period

Flow sign test results for actively managed US mutual funds over the full sample period 1991-2019. The table shows a comparison between model sign and flow sign agreement. The first two columns report the average percent of times that model alpha sign predicts flow sign in the following period, and the statistical significance of this result. The remaining columns show the statistical significance of pairwise tests between the models. Standard errors are double clustered by fund and time.

Flow Sign Test - US Mutual Funds, Active												
	$\frac{\beta+1}{2}$	T-stat	Rating ≥ 4	Rating ≥ 3	CAPM	Mkt-Adj	FF3	FF4	Bmk-Adj	Exc Ret	IR	SR
Rating ≥ 5	68.65%	41.18	6.82	10.59	6.65	16.07	8.48	8.74	19.49	16.67	23.07	25.25
Rating ≥ 4	65.22%	46.13		10.13	11.11	16.57	13.72	14.23	21.06	13.55	25.7	23.84
Rating ≥ 3	62.41%	40.97			5.67	8.97	8.73	9.29	13.20	10.99	17.21	19.70
CAPM	59.68%	28.03				3.40	3.64	4.21	5.65	8.94	10.71	14.69
Mkt-Adj.	59.16%	31.17					1.45	2.11	6.55	7.21	7.91	13.54
FF 3-factor	58.63%	29.42						1.28	2.79	7.78	8.50	13.03
FF 4-factor	58.47%	29.66							2.23	7.50	8.24	12.97
Bmk-Adj.	57.81%	30.59								5.02	4.57	11.02
Excess Return	54.65%	9.04									-2.01	4.73
Info. Ratio	56.07%	16.29										7.30
Sharpe Ratio	51.83%	4.16										

These finding are in line with Ben-David et al. (2020). I also find that the CAPM outperforms the other factor models and market-adjusted, benchmark-adjusted, and excess of risk-free rate returns which is consistent with Berk and van Binsbergen (2016)'s findings.

With regards to how the novel performance measures of Sharpe ratio and Information ratio predict flow sign, both predict sign flow correctly over 50% of the time with statistical significance but not nearly as well as the Morningstar ratings. It is worth noting that for the flow sign test, I have used the Information ratio sign and the Sharpe ratio difference sign in the analysis, where Sharpe ratio difference at time t refers to the difference in Sharpe ratios between time t and t-1. The reason for looking at the sign of the difference in Sharpe ratio from one period to the next rather than simply at the sign of the Sharpe ratio is that otherwise the sign of the Sharpe ratio would simply be the sign of the numerator of its formula (equivalent to the annualized fund return less annualized risk-free rate return since fund inception), and would not include any risk adjustment.

5.2 Style Effects

Next, I test whether investment style affects which asset pricing model investors use in allocating their capital. One hypothesis would be that value investors are possibly more aware of the Fama French factors, since Value is one of the major factors. Consequently, value mutual fund investors could be expected to rely on the Fama-French 3 Factor model and Fama-French-Carhart 4 Factor model more than on the CAPM model or the Morningstar ratings, which do not include a risk adjustment for value. I may also expect growth investors to be less familiar with the multi-factor models than value investors, and to consequently rely on the CAPM and/or Morningstar ratings, which do not include risk adjustment for value, more.

In Table 5 I present a comparison of how well the various performance measures predict sign flow for funds in different investment styles such as Value, Growth, and Blend. For all investment styles and the full sample, I use active US mutual fund data over the entire 1991-2019 period. I acquire mutual fund style information from the "Equity Style Box (Long)" data point in my dataset. My sample contains 124,912 fund-month observations for the Value style, 205,550 fund-month observations for the growth style, and 118,507 fund-month observations for the blend style. The split between styles is hence not too lopsided, and each style contains at least 25% of the total fund-month observations for active mutual funds in my dataset.

Interestingly, I find that Value, Growth, and Blend style investors appear to behave very similarly to the full sample of active mutual fund investors in terms of which asset model they most closely rely on in allocating their capital. For funds in each investment style, flow direction is best predicted by the Morningstar 5 star indicator, followed by the Morningstar 4 star and 3 star indicators.

If Morningstar ratings are not considered, the CAPM alphas would be the strongest predictor of flow direction from the remaining performance measures for each investment style and the overall sample. Value style mutual funds do have slightly higher values of percent agreement between flow sign and the CAPM, 3 Factor, and 4 Factor alphas and a slightly lower value of

Table 5: Comparison of Flow Sign Test Results by Style

This table compares the flow sign test results between all active US mutual funds, Value style active US mutual funds, Growth style active US mutual funds, and Blend style active US mutual funds in my sample. All styles considered include data for the full period sample period 1991-2019. I report the percentage agreement of model alpha signs and flow signs and the statistical significance of these estimates. Standard errors are double clustered by fund and time.

Comparison of Flow Sign Test Results: By Style

	Active, All	**Active, Value**	**Active, Growth**	**Active, Blend**
Rating $\geq$ 5	68.65%	68.12%	69.37%	68.33%
	(41.18)	(23.62)	(29.53)	(20.04)
Rating $\geq$ 4	65.22%	65.00%	65.92%	64.54%
	(46.13)	(24.34)	(34.80)	(23.03)
Rating $\geq$ 3	62.41%	62.75%	62.24%	62.54%
	(40.97)	(23.66)	(29.13)	(21.05)
CAPM	59.68%	60.69%	59.39%	59.48%
	(28.03)	(19.00)	(21.81)	(16.61)
Market-Adjusted	59.16%	60.29%	58.87%	58.96%
	(31.17)	(20.95)	(23.17)	(21.35)
FF 3-factor	58.63%	59.06%	58.72%	58.22%
	(29.42)	(18.55)	(23.16)	(17.34)
FF 4-factor	58.47%	58.88%	58.55%	58.02%
	(29.66)	(18.57)	(24.15)	(16.53)
Benchmark-Adjusted	57.81%	57.24%	58.44%	57.47%
	(30.59)	(17.52)	(26.44)	(18.68)
Excess Return	54.65%	55.87%	53.81%	54.90%
	(9.04)	(9.46)	(6.71)	(8.42)
Information Ratio	56.07%	56.32%	56.21%	56.57%
	(16.29)	(9.79)	(12.24)	(9.87)
Sharpe Ratio	51.83%	52.32%	51.60%	51.71%
	(4.16)	(4.21)	(3.37)	(3.40)

percent agreement between flow sign and the Morningstar 5 star indicator, as predicted by my hypothesis. However, this difference between value style funds and the other style funds and total sample is not very large.

5.3 Management Approach Effects

Next, I test whether the management approach affects which asset pricing model investors prefer. I consider actively managed and passively managed mutual funds, as well as a breakout by time periods for the active mutual fund sample. Since the objective of passive funds is generally to simply track an index, investors would not be expected to invest in passively managed funds with the expectation of outperformance or to direct their flows according to such outperformance. However, surprisingly I find that flows into passive funds do appear to depend on model alphas and especially on Morningstar ratings. Although the magnitude of the flow agreement with Morningstar ratings and model alphas is not as high for passive funds as it is for active funds, it is still significant. This poses an interesting incentive question: if passive funds are rewarded by higher inflows for outperformance, could they benefit from taking on a slightly more active approach while still marketing as a passive fund?

Table 6 presents a comparison of the models for the different management approaches and different time periods. The first column of Table 6 is the same as for the full sample of active mutual funds. Now I have broken up the sample into two time periods, what I refer to as First Period of 1991-2011 and 2nd Period of 2011-2019, so the following two columns of the table refer to actively managed mutual funds during those two respective periods. Finally, the last column of the table considers the passively managed mutual funds over the entire 1991-2019 period.

The rationale behind including multiple time periods is to see whether investor preferences towards certain models have changed over time. The first time period also corresponds to the time period considered in Barber et al. (2016) so I can see the consistency of my results. I find that across the board, model alphas, adjusted returns, Sharpe and Information ratios, and Morningstar ratings explain flows slightly less well in the second time period. The only model that predict flow direction better in the 2nd period than in the 1st period is the 4 Factor model. Consequently, this could indicate that average investor sophistication has increased in recent years.

Table 6: Comparison of Flow Sign Test Results by Time Period and Management Approach

This table compares the flow sign test results between active US mutual funds over the full 1991-2019 time period ("Full Period"), active US mutual funds over the first 1991-2011 time period ("1st Period"), active US mutual funds over the second 2011-2019 time period ("2nd Period"), and passive US mutual funds over the full 1991-2019 time period ("Full Period"). I report the percentage agreement of model alpha signs and flow signs and the statistical significance of these estimates. Standard errors are double clustered by fund and time.

Comparison of Flow Sign Test Results: By Time Period and Management Approach

	Active, Full Period	**Active, 1st Period**	**Active, 2nd Period**	**Passive, Full Period**
Rating $\geq$ 5	68.65%	68.11%	67.78%	62.54%
	(41.18)	(36.56)	(21.16)	(7.29)
Rating $\geq$ 4	65.22%	65.45%	63.57%	56.01%
	(46.13)	(40.53)	(27.70)	(4.17)
Rating $\geq$ 3	62.41%	63.13%	60.67%	55.81%
	(40.97)	(36.61)	(25.65)	(3.90)
CAPM	59.68%	60.32%	58.19%	55.02%
	(28.03)	(22.58)	(19.00)	(5.28)
Market-Adjusted	59.16%	59.05%	57.87%	55.53%
	(31.17)	(23.33)	(22.36)	(7.21)
FF 3-factor	58.63%	58.48%	58.42%	53.59%
	(29.42)	(21.34)	(20.33)	(4.22)
FF 4-factor	58.47%	58.29%	58.31%	53.39%
	(29.66)	(21.00)	(20.79)	(4.14)
Benchmark-Adjusted	57.81%	57.56%	57.46%	54.41%
	(30.59)	(23.82)	(21.73)	(3.95)
Excess Return	54.65%	56.82%	53.72%	51.11%
	(9.04)	(11.45)	(3.93)	(1.24)
Information Ratio	56.07%	57.41%	52.90%	53.61%
	(16.29)	(16.81)	(5.62)	(2.20)
Sharpe Ratio	51.83%	53.19%	50.48%	51.78%
	(4.16)	(5.69)	(0.91)	(2.45)

With regards to passively managed mutual funds, I find that Morningstar rating 5 star, 4 star, and 3 star indicators predict flow sign 62.54%, 56.01%, and 55.81% of the time. All of these measures are significantly higher than the expected 50% at the 1% significance level.

Furthermore, although they do so less well than Morningstar ratings, the CAPM, 3 Factor, and 4 Factor model alphas as well as market-adjusted and benchmark-adjusted returns all predict flow direction to passive funds significantly better than 50% of the time as well.

5.4 Do Flow Signs tell the whole story? Looking at Flow Magnitude

Next, I extend my analysis to look not only at the signs of flows but also at their magnitude. I do so as follows. At each time t, I sort funds into the top-ranked and bottom-ranked buckets according to their Morningstar rating, so that the top-ranked funds all have 5 stars at time t and the bottom-ranked funds all have 1 star at time t. Next, I compute the average percent of positive flows that the top ranked funds received at time t+1, the average percent of positive flows that the bottom ranked funds received at time t+1, and the different between the two. I also look at the percent magnitude of the average flow at time t+1 for the top-ranked funds, for the bottom-ranked funds, and the difference between the two. Finally, I calculate the dollar-value average flow that the top-ranked funds receive at time t+1, the dollar-value average flow that the bottom-ranked funds receive at time t+1, and the difference between the two.

At each point in time when I sort funds into the top-ranked and bottom-ranked funds according to the Morningstar rating, I also sort funds into an equivalent number of top-ranked and bottom-ranked funds according to time t market-adjusted return, benchmark-adjusted return, CAPM alpha, 3 Factor alpha, and 4 Factor alpha. For all of these top-ranked and bottom-ranked funds according to different models, I also compute the average fraction of positive flows next period, the average fund flows as a percent next period, and the average dollar value of fund flows next period in the same way as I did for the Morningstar rating top-ranked and bottom-ranked funds. The results are shown in Table 7.

Table 7 shows that funds in the highest Morningstar category receive a higher percentage of positive flows than funds in the lowest Morningstar category, and that this difference is larger between top-ranked and bottom-ranked Morningstar rated funds than the analagous difference between top-ranked and bottom-ranked funds using any of the other performance measures.

Table 7: **Flows to Top-Ranked and Bottom-Ranked Funds**

This table shows the fraction of positive flows, average amount of flows (in percent), and average amount of flows (in $millions) to top-ranked funds and bottom-ranked funds according to different models. Computed for actively managed US Mutual Funds over the time period 1997-2019. Fund-month observations with missing Morningstar ratings or missing model alphas are excluded from the analysis.

	Fraction Positive Flows			**Fund Flows (%)**			**Fund Flows ($mm)**		
	High	**Low**	**Diff**	**High**	**Low**	**Diff**	**High**	**Low**	**Diff**
Morningstar	67.52%	14.86%	52.66%	1.91%	-1.60%	3.51%	$39.33	-$9.29	$48.62
Market-Adj.	61.50%	14.59%	46.91%	1.89%	-2.13%	4.02%	$22.87	-$24.02	$46.89
Benchmark-Adj.	56.66%	14.83%	41.83%	1.57%	-2.13%	3.70%	$19.56	-$23.89	$43.45
CAPM	61.64%	13.12%	48.52%	1.95%	-2.20%	4.15%	$23.80	-$23.73	$47.53
FF 3-factor	58.60%	12.36%	46.25%	1.77%	-2.15%	3.91%	$19.29	-$24.17	$43.46
FF 4-factor	58.57%	12.92%	45.65%	1.78%	-2.07%	3.86%	$19.93	-$21.88	$41.81

However, what is especially interesting is that the results are less clear-cut for the absolute fund flow and dollar fund flow measures. Here, the top-rated funds according to the CAPM actually receive higher average fund flows than the top-rated funds by Morningstar rating or the other measures. Furthermore, the difference in flows to top-ranked funds and bottom-ranked funds is also highest between the top and bottom-ranked funds according to the CAPM. In terms of dollar flows, the difference between top-rated funds and bottom-rated funds according to Morningstar is similar to that according to the CAPM.

I conclude that while Morningstar might be a better predictor of the signs of flows, the CAPM actually performs comparably well in terms of predicting the average magnitude of flows in percent and dollar terms. It is possible that larger, more sophisticated investors rebalance large amounts according to the CAPM less frequently, while less sophisticated investors rebalance small amounts according to Morningstar ratings more frequently.

In Figure 2, I graph the annual difference between flows to top-ranked and bottom-ranked funds according to the Morningstar rating, CAPM model, and 4 Factor model. I find this difference to be comparable across the three measures of performance, with the most recent data suggesting that more dollar flows were rebalanced according to the CAPM and 4 Factor models than by Morningstar rating.

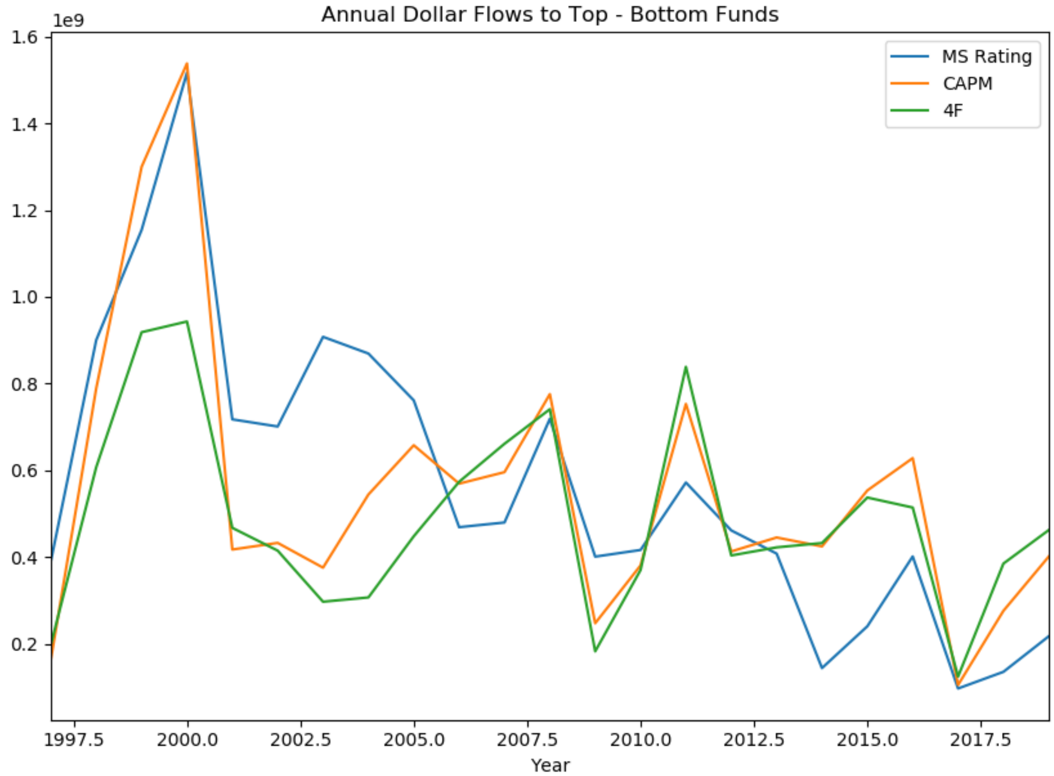


Figure 2: Average Percent Positive Flows to Top-Ranked Funds (by model)

Figure 3: The figure shows a time series of annual difference in dollar flows to top ranked funds by different models and performance metrics. The dataset used is the active US Mutual Fund dataset over 1997-2019.

5.5 Separate Account Composites Performance

I first consider the average performance, relative to different benchmarks and asset pricing models, of the US Separate Accounts Composites dataset. I find that average market-adjusted gross alphas and net alphas for the sample period are positive, 0.12% for gross and 0.04% for net on a monthly basis. On an annual basis, this would translate to 1.4% for gross market-adjusted alpha and 0.5% for net market-adjusted alpha. This is in line with Gerakos et al. (2019)'s recent result, who finds positive gross and net benchmark-adjusted alphas for active institutional accounts over the period 2000-2012. Below I show a plot of average cumulative gross returns, net returns, and S&P 500 returns over the sample period 01/01/1991 - 09/30/2020 by demonstrating how a $1 investment would have grown over time:

Figure 4: Cumulative gross, net, and market returns over the period 1991-2019 for the US Separate Account Compsites - Active. Monthly gross and net returns are computed as the averages of all available gross and monthly fund returns on a given date. Cumulative gross and net returns are then found by cumulating these average monthly returns over time. Market return is taken from Ken French's website.

As can be seen from Figure 4, actively managed US separate accounts have on average outperformed the S&P 500 index over the period 1991-2020. It is important to note, however, that not all of the funds are large-cap or benchmarked to the S&P 500, so some of the outperformance could be due to loadings on factors such as size.

To drill a bit deeper into separate account composite performance, I also present aggregated average alphas for other benchmarks and models. In particular, I include (both gross and net) market-adjusted returns, benchmark-adjustd returns, returns in excess of the risk-free rate, CAPM alphas, Fama-French 3 Factor model alphas, and Fama-French-Carhart 4 Factor model alphas. I present my results in the following table.

From Table 8, I can make several deductions. First, note that while both gross and net average market adjusted return is positive over the entire sample period of 1991-2020, this is largely due to the first portion of the sample period. In fact, while the gross and net market adjusted alphas are on average positive over 1991-2011, they actually become much smaller

Table 8: **Model Alphas. US Composite Accounts, Active**

Average annualized gross and net alphas with respect to the six main models considered in this paper. Market return and risk-free rate come from the Ken French website. Benchmark returns come from Morningstar. Model alphas are averaged across all existing data points for all actively managed US Composites across the period 1991-2020.

Gross and Net Alphas, by model

	1991-2020		1991-2011		2012-2020	
Model Alpha	**Gross**	**Net**	**Gross**	**Net**	**Gross**	**Net**
Market Adjusted Return	1.4%	0.5%	3.1%	2.2%	-0.5%	-1.5%
Benchmark Adjusted Return	1.2%	0.2%	2%	1.1%	-0.4%	-1.3%
Return Excess of Risk-Free Rate	9.2%	8.3%	7.3%	6.4%	1.1%	0.2%
CAPM Alpha	0.3%	-0.6%	2.5%	1.6%	-1.1%	-2%
3F Alpha	0%	-0.9%	0.8%	-0.1%	-0.2%	-1.1%
4F Alpha	0.1%	-0.8%	1%	0.1%	-0.2%	-1.2%

and both gross and net market-adjusted alpha become negative over the most recent time period 2012-2019. Thus while the first portion of the sample (1991-2011) might point towards significant skill of separate account composite managers, the second portion of the sample (2012-2020) appears to refute it. The conclusion I draw is that average alphas for separate account composites appear to have decreased over the most recent decade compared to the previous one. However, even the full sample period of 1991-2020 is not sufficiently long to draw definitive conclusions regarding whether separate account composite managers do or do not have skill.

Another important point to make is that market-adjusted returns, benchmark-adjusted returns, and excess of risk free rate returns exist for almost all fund-month observations, but CAPM alphas, 3 Factor alphas, and 4 Factor alphas do not due to the nature of their construction. As was explained in Section 3, I construct model alphas by first deriving a fund's factor loadings using rolling 60 months observations or fund returns and factor returns. Then I derive the model alphas for a fund in a given month by subtracting from that month's fund return the product of current factor returns and the previous month's calculated fund factor loadings. Consequently, I do not have model alphas for the first 60 months of a fund's life. This is important to note when looking at the low factor model alphas for the entire sample period 1991-2019 in Table

8, especially when compared to higher market-adjusted, benchmark-adjusted, and excess of risk-free rate returns. This is predominantly due to the fact that the averages are taken over existing data point values, and factor model alphas only exist in the latter portion of that period. Nevertheless the returns for separate account appear to be on average higher than those for mutual funds over the same time period, even on a risk-adjusted basis.

5.6 Flows Sign Test for Separate Account Composites

Next, I repeat the flow sign test on the US Separate Account Composites dataset. Through this test, I am aiming to see if separate account investors respond differently to model alphas and Morningstar ratings than retail investors. In particular, I am interested in seeing whether they tend to use more sophisticated models such as the multi-factor models rather than simple ratings such as the Morningstar Rating.

Table 9: Flow Sign Test - US Separate Account Composites, Active, Full Period

Flow sign test results for actively managed US separate account composites over the full sample period 01/01/1991 - 09/30/2020. The table shows a comparison between model sign and flow sign agreement. The first two columns report the average percent of times that model alpha sign predicts flow sign in the following period, and the statistical significance of this result. The remaining columns show the statistical significance of pairwise tests between the models. Standard errors are double clustered by fund and time.

Flow Sign Test - US Separate Account Composites, Active												
	$\frac{\beta+1}{2}$	T-stat	Rating ≥ 4	Rating ≥ 3	FF3	FF4	CAPM	Bmk-Adj	Mkt-Adj	IR	Exc Ret	SR
Rating ≥ 5	62.26%	19.02	6.78	8.86	11.71	11.84	12.08	12.74	12.83	12.05	13.64	5.76
Rating ≥ 4	57.94%	19.47		6.00	9.37	9.38	9.76	10.11	10.61	9.18	10.31	4.58
Rating ≥ 3	55.64%	13.66			2.83	2.95	3.59	4.58	4.85	5.55	7.95	3.71
FF 3-factor	54.51%	16.02				0.62	2.02	3.64	4.69	3.80	7.69	3.01
FF 4-factor	54.44%	15.64					1.59	3.19	4.26	3.69	7.48	2.98
CAPM	54.07%	12.69						1.28	3.48	2.87	7.12	2.71
Bmk-Adj.	53.67%	14.07							0.90	2.37	6.31	2.46
Mkt-Adj.	53.45%	11.62								1.85	6.20	2.31
Info. Ratio	52.43%	4.82									2.47	1.53
Excess Return	50.78%	1.67										0.17
Sharpe Ratio	50.58%	0.51										

As Table 9 shows, the Morningstar rating 5 star indicator is also the best predictor of sign flow for US Separate Account Composites. Flow sign is predicted correctly by this metric

62.26% of the time and is statistically significant at the 1% level. Within factor models, the Fama-French 3 Factor model and Fama-French-Carhart 4 Factor model predict flow sign better than the CAPM.

In addition to the flow sign test above, I also perform a flow sign test of the Information Ratio forms of the model alphas. I compute these Information Ratio forms of the alphas as follows. First, for a given model I calculate the cumulative model alpha since inception. Next, I calculate the annualized standard deviation of the monthly model alphas since inception. I then divide the cumulative model alpha return by the annualized standard deviation of the model alphas, which essentially computes a since inception information ratio using that model. Finally, I compute a weighted value of the past 18 months of these information ratios as I computed for regular model alphas in the flow sign test earlier. This is in accordance with Barber et al. (2016)'s methodology and the reason for it is that investor's may have a slightly delayed reaction to the most recent performance measures.

The results of the information ratio flow sign test are summarized in Table 10, and show that in this more comparable setting the Fama-French 3 Factor model and Fama-French-Carhart 4 Factor model information ratios predict flows next period with the highest degree of magnitude and statistical significance. In this more comparable setting, I also not that the regular information ratio - computed relative to the specified composite benchmark - actually outperforms the CAPM information ratio, indicating that performance ratios such as the information ratio may be viewed by sophisticated investors to be as or even more important than performance relative to the CAPM. The difference in predictability for flows using the information ratio and CAPM information ratio is not statistically significant, however.

5.7 Panel Regression Test

Next, I extend the analysis by looking at what variables are most significant in explaining flows in a regression context. I include prior flows, log fund size, and log fund age as control variables.

Table 10: Information Ratio Flow Sign Test - US Separate Account Composites, Active, Full Period

Information ratio flow sign test results for actively managed US separate account composites over the full sample period 01/01/1991 - 09/30/2020. The table shows a comparison between model sign and flow sign agreement. The first two columns report the average percent of times that model alpha sign predicts flow sign in the following period, and the statistical significance of this result. The remaining columns show the statistical significance of pairwise tests between the models. Standard errors are double clustered by fund and time.

Flow Sign Test - US Separate Account Composites, Active							
	$\frac{\beta+1}{2}$	T-stat	Information Ratio 4F	Information Ratio	Information Ratio CAPM	Information Ratio Mkt	Sharpe Ratio
Information Ratio 3F	54.68%	11.00	1.41	4.14	5.5	4.54	3.13
Information Ratio 4F	54.28%	10.47		3.4	4.88	4.01	2.88
Information Ratio	52.43%	4.82			0.27	0.45	1.53
Information Ratio CAPM	52.29%	5.18				0.09	1.46
Information Ratio Mkt	52.25%	4.77					1.44
Sharpe Ratio	50.58%	0.51					

First, I compare the CAPM model, 3 Factor model, 4 Factor model, market-adjusted, benchmark-adjusted, and excess of risk-free rate returns as potential benchmarks that investors use for allocating their flows. I run a panel regression of monthly flows on the various model alphas. I use the actively managed US mutual fund dataset with 463,937 fund-month observations. I include month fixed effects in all regressions to account for market conditions and changes in flow magnitude due to different economic environments. I also include controls for the previous flows in months t-1 to t-18, previous month log fund size, and previous month log fund age relying on prior literature that has shown those to be significant drivers of flows. The results of my regressions are shown in 11.

I find that within factor models, the CAPM appears to perform the best and maintains the highest level of significance when the CAPM, 3 Factor, and 4 Factor alphas are all included as explanatory variables for flows. When I add in market-adjusted, benchmark-adjusted, and excess returns the CAPM still holds the highest level of significance. In general, the factor model alphas have higher significance than the adjusted return performance measures. Within the adjusted return measures, benchmark-adjusted returns maintain higher significance than market-adjusted returns or excess of risk-free rate returns; in a regression including all six performance measures, only the factor models and benchmark-adjusted return remain significant at the 1% level.

Table 11: Panel Regression - US Mutual Funds - Factor Models

Panel regressions of fund flows on previous month weighted CAPM alphas, weighted 3 Factor alphas, weighted 4 Factor alphas, weighted market-adjusted alphas, weighted benchmark-adjusted alphas, and weighted excess or risk-free rate alphas. All regressions include time-fixed effects. Standard errors are double clustered by fund and time. The control variables include the fund's prior flows for months t-1 through t-18, log of previous month fund size, and log of previous month fund age. ***, **, and * denote significance at the 10%, 5%, and 1% level respectively.

Panel Regression - US Mutual Funds, Active - Factor Models								
	Flow (1)	Flow (2)	Flow (3)	Flow (4)	Flow (5)	Flow (6)	Flow (7)	Flow (8)
CAPM Alpha	1.03***						0.63***	0.54***
	(19.46)						(12.28)	(6.09)
3 Factor Alpha		1.26***					0.41***	0.28***
		(20.65)					(4.46)	(3.12)
4 Factor Alpha			1.26***				0.33***	0.29***
			(20.10)				(3.75)	(3.18)
Market-Adjusted Return				1.06***				1.43*
				(15.77)				(1.66)
Benchmark-Adjusted Return					1.17***			0.31***
					(16.01)			(5.92)
Excess Return						0.97***		-1.41
						(15.65)		(-1.64)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	463,937	463,937	463,937	463,937	463,937	463,937	463,937	463,937
R^2	2.09%	2.05%	2.01%	2.51%	2.44%	2.43%	2.18%	2.24%

Next, I add performance metrics such as Morningstar rating, Sharpe ratio, and Information ratio to my regression. The results are shown in Table 12. As before, control variables include prior flows, fund size, and fund age. All regressions include month fixed effects. Standard errors are double clustered by fund and time. The dataset used is the actively managed US mutual funds.

Now I find that the Morningstar rating has higher significance than any of the factor models or adjusted return measures. When Morningstar rating, Sharpe ratio, and Information ratio are all included as explanatory variables, Morningstar rating has the highest statistical significance in explaining flows. In fact, Sharpe ratio and Information ratio coefficients become negative, though this may be due to collinearity issues with the Excess return and Benchmark-adjusted return performance measures, respectively. When the CAPM, 3 Factor, 4 Factor models and the various other risk-adjusted performance metrics are added to the regression, the Morningstar rating maintains the highest level of significance.

Table 12: Panel Regression - US Mutual Funds - Morningstar Rating, Sharpe Ratio, Information Ratio

Panel regressions of fund flows on previous month Morningstar rating, Sharpe ratio, Information ratio, weighted CAPM alphas, weighted 3 Factor alphas, weighted 4 Factor alphas, weighted market-adjusted alphas, weighted benchmark-adjusted alphas, and weighted excess or risk-free rate alphas. All regressions include time-fixed effects. Standard errors are double clustered by fund and time. The control variables include the fund's prior flows for months t-1 through t-18, log of previous month fund size, and log of previous month fund age. ***, **, and * denote significance at the 10%, 5%, and 1% level respectively.

Panel Regression - US Mutual Funds, Active - Performance Metrics					
	Flow (1)	Flow (2)	Flow (3)	Flow (4)	Flow (5)
Morningstar Rating	0.0090*** (26.87)			0.0080*** (23.15)	0.0063*** (16.10)
Sharpe Ratio		0.0102*** (13.04)		0.0031*** (4.09)	-0.0057*** (-4.74)
Information Ratio			0.0088*** (13.90)	0.0031*** (5.76)	-0.0018*** (-3.47)
CAPM Alpha					0.40*** (4.60)
3 Factor Alpha					0.19** (2.11)
4 Factor Alpha					0.18** (2.05)
Market-Adjusted Return					1.46* (1.67)
Benchmark-Adjusted Return					0.25*** (5.14)
Excess Return					-1.37 (-1.56)
Month FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of Observations	463,937	463,937	463,937	463,937	463,937
R^2	2.46%	1.98%	2.06%	2.50%	2.56%

Next, I look at how these panel regression results change for the separate account composites dataset. I find that on their own, all performance measures considered predict flows at the 1% statistical level. Table 13 column (7) shows that in a comparison of the three factor models, the 4 Factor model now has a higher magnitude and higher statistical significance in predicting flows than the CAPM or 3 Factor models, indicating reliance on the most sophisticated factor model by separate account investors. When all risk-adjusted metrics are included, the CAPM

Table 13: Panel Regression - US Separate Account Composites - Factor Models

Panel regressions of fund flows on previous month weighted CAPM alphas, weighted 3 Factor alphas, weighted 4 Factor alphas, weighted market-adjusted alphas, weighted benchmark-adjusted alphas, and weighted excess or risk-free rate alphas. All regressions include time-fixed effects and Morningstar Category style fixed effects. Standard errors are double clustered by fund and time. The control variables include the fund's prior flows for months t-1 through t-18, log of previous month fund size, and log of previous month fund age. ***, **, and * denote significance at the 10%, 5%, and 1% level respectively.

Panel Regression - US Separate Account Composites - Factor Models								
	Flow (1)	Flow (2)	Flow (3)	Flow (4)	Flow (5)	Flow (6)	Flow (7)	Flow (8)
CAPM Alpha	0.86***						0.28**	0.96***
	(8.22)						(2.19)	(4.34)
3 Factor Alpha		1.30***					0.03	-0.42
		(10.08)					(0.08)	(-1.12)
4 Factor Alpha			1.38***				0.93***	1.08***
			(10.50)				(2.81)	(2.72)
Market-Adjusted Return				0.75***				1.22
				(7.95)				(1.00)
Benchmark-Adjusted Return					1.17***			0.77***
					(11.47)			(4.80)
Excess Return						0.74		-2.08*
						(7.85)		(-1.69)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Style FE	Yes	Yes	Yes	Yes	Yes			
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	103,671	103,671	103,671	130,116	129,783	130,116	103,671	103,510
R^2	3.70%	3.72%	3.73%	3.13%	3.17%	3.12%	3.74%	3.76%

alpha, Fama-French 4 Factor model alpha, and Benchmark-Adjusted Return maintain a 1% level of significance.

When Morningstar ratings, Sharpe ratio, and Information ratio are added to the regression in Table 14, the Morningstar rating, Information Ratio and Sharpe Ratio are the only metrics that remains statistically significant at the 1% level in all regressions. This indicates that even for separate account composites, Morningstar rating plays a very large role in how investors allocate their capital. It also indicates that separate account investors pay attention to benchmarking fund performance, and considering performance relative to factor models is unlikely to be the only metric of performance evaluation that they use. This lends some support to Beber et al. (2019)'s conclusion that investment funds should be judged using bespoke benchmarks - which include the fund's specific leverage constraints, shorting constraints, position limits and so forth

Table 14: Panel Regression - US Separate Account Composites - Morningstar Rating, Sharpe Ratio, Information Ratio

Panel regressions of fund flows on previous month Morningstar rating, Sharpe ratio, Information ratio, weighted CAPM alphas, weighted 3 Factor alphas, weighted 4 Factor alphas, weighted market-adjusted alphas, weighted benchmark-adjusted alphas, and weighted excess or risk-free rate alphas. All regressions include time-fixed effects and Morningstar Category style fixed effects. Standard errors are double clustered by fund and time. The control variables include the fund's prior flows for months t-1 through t-18, log of previous month fund size, and log of previous month fund age. ***, **, and * denote significance at the 10%, 5%, and 1% level respectively.

Panel Regression - US Separate Account Composites, Active - Performance Metrics					
	Flow (1)	Flow (2)	Flow (3)	Flow (4)	Flow (5)
Morningstar Rating	0.0112***			0.0100***	0.0085***
	(9.44)			(8.98)	(5.82)
Sharpe Ratio		0.0028		0.0019*	0.0035*
		(1.52)		(1.66)	(1.70)
Information Ratio			0.0117***	0.0104***	0.0080***
			(6.99)	(4.39)	(3.19)
CAPM Alpha					0.30
					(1.08)
3 Factor Alpha					-0.31
					(-0.76)
4 Factor Alpha					0.61
					(1.61)
Market-Adjusted Return					10.25*
					(1.69)
Benchmark-Adjusted Return					0.70***
					(3.73)
Excess Return					-10.51*
					(-1.73)
Month FE	Yes	Yes	Yes	Yes	Yes
Style FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of Observations	92,715	130,041	129,837	92,666	81,841
R^2	4.04%	3.06%	3.13%	4.07%	4.62%

- rather than by computing factor model alphas without taking into account such constraints.

6 Tests of the Robustness of my Results

6.1 Splits by Time Period

From the above evidence and discussion, I have seen that the Morningstar rating appears to have a stronger explanatory effect on flows than either the CAPM, 3 Factor, or 4 Factor alphas or Market-adjusted, Benchmark-adjusted, or Excess Returns. Given the relatively novel nature of the Morningstar rating (as mentioned, it was established in 1985), it is a natural question to ask whether the important of Morningstar ratings for investor flows has increased, decreased, or remained constant over time. Have investors gained confidence in the performance rating and started allocating more of their flows in accordance with it? Or on the contrary, have investors realized that the Morningstar rating is not a sufficiently sophisticated and relevant measure for them and started allocating flows according to it less? In Table 6, I have looked at how investor model preferences have changed over time.

6.2 Previous Month Alphas vs Weighted Alphas

Throughout my analysis, I have followed Barber et al. (2016)'s methodology and used weighted alphas in my calculations and results. The rationale behind weighted alphas for investors make sense; more recent fund performance gets higher weight, but investors also consider prior months' fund returns in addition to the single prior month alpha metric. What happens when I simplify the model alpha calculations to include simple prior month model alphas rather than weighted alphas? I have performed robustness checks on such model alphas, and find that their explanatory power for flows is dramatically less than that of weighted alphas. It appears that investors do consider funds' historical returns in addition to return in any given month. Results for this check are included in the Appendix.

6.3 Consecutive Monthly Net Asset Observations

In another robustness check for my separate account composite dataset results, I re-run my analysis using only data for which I have consecutive monthly net asset value data points. The motivating reason for this exercise is that many funds would have consecutive monthly net asset observations for a period of time and then switch to reporting only quarterly, followed by another period of monthly observations. For simplicity in this exercise, I kept only funds that for their entire history provided consecutive (i.e. no skipped month) monthly net asset value observations. The dataset constructed this way had cleaner observation data, but the obvious drawback was the significant reduction in the dataset's size. As can be seen from the descriptive statistics table in Appendix A, the number of fund-month observation then drops to 46,354 for a total of 395 composites each with approximately 10 years of monthly data.

6.4 Datapoint Availability Consistency

A final robustness check I perform relates to the sign flow tests and the fact that various model alphas and Morningstar ratings are available for different funds at different periods. As has been discussed, factor model alphas are calculated using the fund factor loadings, that are themselves calculated from rolling 60 month regressions of fund returns on factor returns. Consequently, for any fund factor model alpha data points will be missing for the first 60 months. For weighted alphas, which are computed using the past 18 month lags, the factor model alpha availability issue is compounded further.

In order to help resolve concerns that my data results are significantly influenced by the availability of alphas and Morningstar ratings for funds at different time periods, I perform the flow sign analysis on US Mutual Funds and US Separate Account Composites now using only datasets that have non-missing observations for all alphas and Morningstar ratings. That is to say, if a fund F has a Morningstar rating in month t but no CAPM Weighted Alpha data point in month t, then the month t, Fund F observation will be dropped from this robustness analysis.

In this way I only look at fund, time observations that have non-missing values for each model (where factor, market-adjusted, or Morningstar rating heuristic) that I want to compare. The results are shown in the Appendix, and are in line with my main results.

7 Prospective Theoretical Explanations

7.1 Bounded Rationality

I next propose several possible explanations for the observed results, to be developed more fully in future research. Why are investor flows driven so strongly by Morningstar ratings rather than more sophisticated factor models? One possible explanation could be bounded rationality. Bounded rationality is the the idea that decision making, for investors as any other human beings, is fraught with practical limitations and difficulties. For one, individuals are limited by the information they have. If investors are not provided with many other key data points about a fund besides the Morningstar rating, for instance, they would be more likely to rely heavily on the Morningstar rating.

On a related note, investors also have a finite amount of time to make their investing decisions. Hence while it is certainly possible for investors to calculate a fund's CAPM Alpha, 3 Factor Alpha, 4 Factor alpha, and other more sophisticated measures of a fund's historical risk-adjusted performance, such computations can be time intensive and not worth the effort for investors. This becomes a doubly strong point for smaller less sophisticated investors, as the effort required to perform more sophisticated calculations is likely higher (potentially requiring additional financial education), and as the notional invested is smaller the impact of assessing fund performance is less in dollar-value terms.

Finally, another limitation brought up by bounded rationality is that even with the available information investors are given, the cognitive limitations of their minds may be prohibiting them from processing all of it. For instance, in today's day and age fund prospectuses include a slew of different fee types, gross and net returns over a number of different time periods and year combinations, maximum drawdowns, worst and best performing years, styles, manager and firm information, and many additional potentially relevant data points. Processing all of this information, and especially deciphering for oneself what is truly relevant for a fund's future performance, can be a daunting task that may not be able to be adequately analyzed by an individual investor. Even within well known and well studied factor models, this issue is compounded by the common behavioural finance problem of choice overload: when faced with so many different performance models, which one should investors use? What happens when the models give different guidance for which funds are the best risk-adjusted performers? Choi and Robertson (2020) make headway in this direction by looking at how real investors make their investment decisions, as they say "straight from the horse's mouth." As can be expected from the discussion in this section, investors tend to make some simplifications when evaluating funds: Choi and Robertson (2020) find that generally investors tend to believe that past mutual fund performance is a good signal of stock-picking skill, actively managed funds do not suffer from diseconomies of scale, value stocks are safer and do not have higher expected returns, and high-momentum stocks are riskier and do have higher expected returns. Roussanov et al. (2020) also consider how mutual fund flows can best be explained in a context of imperfectly rational investors. When estimating a structural model of investor beliefs and comparing it with the rational Bayesian benchmark based on past performance, they find that investors are more optimistic about fund managers' average skill than warranted by the historical data, over-weight recent performance in a manner consistent with models based on the "representativeness" heuristic, and respond slowly to changes in their beliefs which is consistent with bounded rationality issues such as limited attention and informational frictions.

In a trade-off between effort exertion, available time, available information and limitations on information processing, investors (especially smaller less sophisticated ones) may be tempted to seek a magical, all-inclusive measure of fund quality and expected future performance. From what I have seen time and again in my results according to different tests, it appears that for many such a measure is precisely the Morningstar rating.

7.2 The Power of Morningstar Ratings?

Bounded rationality considerations aside, what if the Morningstar rating actually is an extremely good predictor of future fund performance? What if historically, the Morningstar Rating actually has higher power in predicting future fund performance than more sophisticated measures such as the factor model alphas? After all, considerable time and consideration went into the creation of the Morningstar Ratings by financial investment professionals; furthermore, the ratings look at both risk-adjusted performance and consider funds in the larger context of portfolio allocations. Consequently, one possible reason, so far untested, for why investors rely so heavily on Morningstar ratings in allocating their capital is perhaps because those ratings actually work the best in predicting future performance. To be able to reject this hypothesis and move onto the bounded rationality concerns discussed earlier, I need to look at historical performance and compare the return predictability for CAPM alphas, 3 Factor alphas, and 4 Factor alphas with the return predictability for Morningstar ratings. Insofar as I am aware this has not been backtested in the academic literature, and would constitute an interesting and novel result. This is one of the imminent extensions I am working on for this paper.

7.3 Frictions and Rigidity

Most investors in mutual funds are retail investors who invest through their retirement plans. This could be a potential mechanism that leads investors to place such high value on Morningstar ratings. One supposition is that because the majority of mutual fund assets come from people's retirement accounts, it is the marketing and availability of retirement investment options that plays a large role in determining where investors place their money. There are at least two potential mechanisms here: 1) first, investors might have limited availability for the funds that they are allowed to invest in through their pension plan, and 2) investors might be shown limited fund performance summary statistics, which would often include Morningstar rating but not such metrics as the 3 Factor alpha or 4 Factor alpha.

8 Conclusion

To conclude, I summarize my main findings and discuss future areas of potential research. In this paper, I have assessed whether Morningstar ratings do indeed overshadow other model alphas in explaining investor flows, and find that they generally do; however, the effect is less strong for separate account composites, that generally include more sophisticated investors, than for mutual funds. Surprisingly, though the effect is dampened, Morningstar ratings still matter even for those mutual funds running a passive strategy. By breaking up my data sample by time period and style, I also see how much the Morningstar ratings drive flows for funds in different time periods (1991-2011 vs 2012-2019), with different management approaches (active vs passive) and different style (value vs growth vs blend).

In this paper I have dug deeper into the Morningstar rating influence on flows phenomenon by considering different time periods, differnet management approaches, different investment styles, and additional risk-adjusted performance metrics within the context of active mutual funds, passive mutual funds, and separate account composites. I also included two other similarly widespread easy-to-use metrics of risk-adjusted performance, the Sharpe ratio and the Information Ratio, and saw how those stack up against the Morningstar rating and factor model alphas in terms of explaining investor flows.

These empirical findings shed some light on how investors allocate their capital, but they also raise many other questions. For instance, what is the theoretical groundwork for why Morningstar ratings matter to investors? In other words, what is the channel through which investors receive this information, do they seek it out, and why does it appear to supercede other performance-related information investors have about the fund?

Another natural question also arises relating to a feedback loop where, knowing investors care deeply about Morningstar ratings, fund managers might attempt to maximize them at the expense of the long-term best performance of the fund. For instance, Morningstar ratings are calculated for the ending 3 years, 5 years, and 10 year periods; so would fund managers exert more effort when their fund reaches their first 5 years or first 10 years, as those new ratings will then appear? The literature on earnings management is quite extensive. Agarwal et al. (2011), for instance, show earnings management for hedge funds. It would be interesting to see if a similar thing occurs for mutual funds with regards to maximizing Morningstar ratings. Since the ratings are given within Morningstar Category, would funds also potentially attempt to reclassify themselves into a different Morningstar Category where they would rank more highly (and receive a higher Morningstar Rating) among competitors? These and other related questions are an interesting direction of further research, which I hope to address in the future.

Chevalier and Ellison (1997) show that while investors would like mutual funds to maximize risk-adjusted fund returns, fund managers' incentives are not only to maximize performance but also to maximize flows. They demonstrated that mutual fund managers alter the riskiness of their portfolios near the end of the year dependent on the fund's year-to-date return. Consequently, another action that increases flows - such as obtaining a higher Morningstar rating - would be seen as favorable by fund managers and could lead to suboptimal (from the perspective of investors) portfolio management decisions. Since I have seen that even passive funds' flows are guided by outperformance and Morningstar rating metrics, these incentives could extend to passive funds as well.

Another possible future direction would be to extend the analysis to other investment vehicles such as hedge funds. Is there an equivalent metric to Morningstar ratings for hedge funds, and if so does it have a significant impact on investor flows? Or do hedge fund investors, due to their more sophisticated nature, avoid allocating capital according to simple fund ratings? In the hedge fund industry, just as for mutual funds, fund managers have also been shown to act towards incentives misaligned with those of investors and it would be interesting to see if there is a rating measure that exacerbates this issue.

References

Agarwal, V., N. D. Daniel, and N. Y. Naik (2011): "Do Hedge Funds Manage Their Reported Returns?" *Review of Financial Studies 24 (10): 3181-3320.*

Barber, B., X. Huang, and T. Odean (2016): "Which Factors Matter to Investors? Evidence from Mutual Fund Flows," *Review of Financial Studies 29 (10): 2600-2642.*

Beber, A., M. Brandt, J. Cen, and K. Kavajecz (2019): "Mutual Fund Performance: Using Bespoke Benchmarks to Disentangle Mandates, Constraints and Skill," *Working Paper.*

Ben-David, I., J. Li, A. Rossi, and Y. Song (2020): "What Do Mutual Fund Investors Really Care About?" *Working Paper.*

Berk, J. B. and R. C. Green (2004): "Mutual Fund Flows and Performance in Rational Markets," *Journal of Political Economy 112 (6) 1269-1295.*

Berk, J. B. and J. H. van Binsbergen (2016): "Assessing Asset Pricing Models using Revealed Preference," *Journal of Financial Economics 119 (1): 1-23.*

Carhart, M. M. (1997): "On Persistence in Mutual Fund Performance," *The Journal of Finance 52 (1): 57-82.*

Chevalier, J. and G. Ellison (1997): "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economics 105 (6): 1167-1200.*

Choi, J. J. and A. Z. Robertson (2020): "What Matters to Individual Investors? Evidence from the Horse's Mouth," *The Journal of Finance Forthcoming.*

Cremers, M. and A. Petajisto (2009): "How Active is Your Fund Manager? A New Measure That Predicts Performance," *Review of Financial Studies 22 (9): 3329-3365.*

Del Guercio, D. and P. A. Tkac (2008): "Star Power: The Effect of Morningstar Ratings on Mutual Fund Flow," *The Journal of Financial and Quantitative Analysis 43 (4) 907-936.*

Evans, R. B. and R. Fahlenbrach (2012): "Institutional Investors and Mutual Fund Governance: Evidence from Retail-Institutional Fund Twins," *Review of Financial Studies 25 (12): 3530-3571.*

Evans, R. B. and Y. Sun (2020): "Models or Stars: The Role of Asset Pricing Models and Heuristics in Investor Risk Adjustment," *The Review of Financial Studies Forthcoming.*

Fama, E. F. and K. R. French (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics 33 (1): 3-56.*

——— (2015): "A five-factor asset pricing model," *Journal of Financial Economics 116 (1) 1-22.*

Gerakos, J., J. T. Linnainmaa, and A. Morse (2019): "Asset Manager Funds," *Working Paper.*

Jagannathan, R. and Z. Wang (1996): "The CAPM is Alive and Well," *The Journal of Finance 51 (1): 3-53.*

Jegadeesh, N. and S. Titman (1993): "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *The Journal of Finance 48 (1) 65-91.*

Pastor, L. and R. Stambaugh (2002): "Mutual fund performance and seemingly unrelated assets," *Journal of Financial Economics 63 (3) 315-349.*

Pastor, L., R. F. Stambaugh, and L. A. Taylor (2015): "Scale and Skill in Active Management," *Journal of Financial Economics 116 (1): 23-45.*

Roussanov, N., H. Ruan, and Y. Wei (2020): "Mutual Fund Flows and Performance in (Imperfectly) Rational Markets," *Working Paper.*

Sharpe, W. F. (1964): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *The Journal of Finance 19 (3) 425-442.*

——— (1966): "Mutual Fund Performance," *The Journal of Business 39 (1): 119-138.*

Appendix A. Additional Results

Below I provide additional tests and results for the data sample. I include both descriptive results of the datasets and robustness regression and flow sign bivariate regression test results.

Table 15: Descriptive Statistics. US Composite Accounts - Strictly Consecutive Monthly Observations

Descriptive statistics of the observation data sample for US composite accounts, sorted into buckets by Morningstar Rating. This table summarizes the main descriptive statistics for the US Separate Account Composites sample of active funds over the time period January 1991 - December 2019. Observations refers to fund-month observations, and all statistics are computed as averages over these observations.

	Morningstar Rating						
	1 Star	2 Stars	3 Stars	4 Stars	5 Stars	Rating NA	All
Fund-month observations	269	1,380	3,030	2,398	778	38,499	46,354
Fund size ($million)	290.16	999.59	1,394.97	1,698.66	1,717.30	1,123.99	1,172.85
Fund age (years)	9.96	11.02	11.78	10.68	7.31	7.65	8.18
Fund flow	-3.14%	-1.64%	-0.65%	0.34%	3.25%	1.47%	1.18%
Mkt-Adj Return	-0.15%	-0.09%	0.10%	0.19%	0.26%	0.08%	0.08%
Excess Return	0.88%	0.54%	0.82%	0.79%	0.79%	0.83%	0.82%
Ret. Volatility (1yr)	4.63%	4.45%	4.24%	3.97%	3.90%	4.33%	4.30%
Ret. Volatility (5yr)	5.17%	4.89%	4.58%	4.32%	4.03%	4.65%	4.63%
Market beta	0.97	0.99	0.98	0.96	0.94	0.97	0.97
Size beta	0.28	0.26	0.25	0.22	0.19	0.21	0.22
Value beta	0.02	0.11	0.09	0.06	0.05	0.09	0.09
Momentum beta	-0.02	-0.02	0.00	0.01	0.02	0.00	0.00
Fraction of positive flows	21.2%	20.6%	29.1%	39.5%	59.1%	41.16%	39.86%

Table 16: Descriptive Statistics. US Mutual Funds - Strictly Consecutive Monthly Observations

Descriptive statistics of the observation data sample for US Mutual Fund accounts, sorted into buckets by Morningstar Rating. This table summarizes the main descriptive statistics for the US Mutual Funds sample of active funds over the time period January 1991 - December 2019. Observations refers to fund-month observations, and all statistics are computed as averages over these observations.

	Morningstar Rating						
	1 Star	2 Stars	3 Stars	4 Stars	5 Stars	Rating NA	All
Fund-month observations	15,219	59,950	108,540	79,505	29,969	41,128	334,311
Fund size ($million)	600.95	953.87	1848.32	2914.61	3878.17	335.41	1880.56
Fund age (years)	10.66	11.25	11.2	10.62	8.86	1.36	9.62
Fund flow	-1.78%	-1.25%	-0.49%	0.75%	2.95%	5.00%	5.90%
Mkt-Adj Return	-0.39%	-0.12%	0.02%	0.18%	0.52%	0.11%	0.07%
Excess Return	0.23%	0.56%	0.71%	0.86%	1.06%	0.71%	0.73%
Ret. Volatility (1yr)	5.19%	4.63%	4.32%	4.27%	4.57%	4.43%	4.44%
Ret. Volatility (5yr)	5.65%	5.04%	4.71%	4.57%	4.56%	5.22%	4.78%
Market beta	1.04	1.01	0.99	0.97	0.96	0.99	0.99
Size beta	0.29	0.22	0.19	0.18	0.2	0.22	0.2
Value beta	0	0.03	0.05	0.07	0.05	0.03	0.05
Momentum beta	0	0.01	0.01	0.01	0.01	0.01	0.01
Fraction of positive flows	15.24%	19.28%	29.21%	48.96%	71.75%	68.79%	40.17%

Table 17: Flow Sign Test - US Separate Account Composites, Active, Full Period - Strictly Consecutive Monthly Observations

Flow sign test results for actively managed US separate account composites using only composites with strictly consecutive monthly data over the full sample period 1991-2019. The table shows a comparison between model sign and flow sign agreement. The first two columns report the average percent of times that model alpha sign predicts flow sign in the following period, and the statistical significance of this result. The remaining columns show the statistical significance of pairwise tests between the models. Standard errors are double clustered by fund and time.

Flow Sign Test - US Separate Account Composites												
	$\frac{\beta+1}{2}$	T-stat	Rating ≥ 4	Rating ≥ 3	CAPM	Mkt-Adj	FF3	FF4	Bmk-Adj	Exc Ret	IR	SR
Rating ≥ 5	64.33%	11.83	2.42	2.96	2.16	5.21	2.42	2.8	6.07	5.71	6.26	8.09
Rating ≥ 4	61.04%	9.79		1.79	3.82	4.89	4.21	4.72	5.41	5.27	5.6	7.65
Rating ≥ 3	58.54%	6.34			2.08	2.67	2.41	2.69	3.38	3.85	4.35	5.57
CAPM	55.20%	6.78				1.23	1.52	1.44	2.25	5.31	3.04	6.22
Mkt-Adj.	55.10%	8.56					0.52	0.5	2.26	4.19	2.01	7.35
FF 3-factor	54.36%	7.17						-0.14	1.57	4.62	2.61	5.55
FF 4-factor	54.40%	7.37							1.66	4.6	2.63	5.71
Bmk-Adj.	54.03%	8.09								2.78	0.97	5.81
Excess Return	51.79%	2.68									-1.32	2.70
Info. Ratio	53.08%	3.55										3.14
Sharpe Ratio	50.19%	0.42										

Table 18: Flow Sign Test - US Separate Account Composites, Active, 1991-2011

Flow sign test results for actively managed US separate account composites using only composites with strictly consecutive monthly data over the full sample period 1991-2019. The table shows a comparison between model sign and flow sign agreement. The first two columns report the average percent of times that model alpha sign predicts flow sign in the following period, and the statistical significance of this result. The remaining columns show the statistical significance of pairwise tests between the models. Standard errors are double clustered by fund and time.

Flow Sign Test - US Separate Account Composites - 1991-2011

	$\frac{\beta+1}{2}$	T-stat	Rating ≥ 4	Rating ≥ 3	CAPM	Mkt-Adj	FF3	FF4	Bmk-Adj	Exc Ret	IR	SR
Rating ≥ 5	60.91%	12.96	2.8	3.84	2.53	5.15	2.64	3.18	5.02	1.47	8.63	2.27
Rating ≥ 4	58.90%	15.39		3.08	3.59	5.39	2.98	3.71	5.52	1.05	11.01	1.78
Rating ≥ 3	57.53%	13.24			-0.17	3.8	-0.17	0.56	3.50	0.68	8.84	1.46
CAPM	54.48%	6.38				1.70	2.57	1.80	1.53	3.73	2.00	3.25
Mkt-Adj.	53.74%	6.90					1.36	0.69	-0.91	1.31	0.87	1.20
FF 3-factor	52.99%	4.92						-1.36	-1.04	2.00	0.66	1.69
FF 4-factor	53.42%	5.78							-0.17	2.49	1.09	2.17
Bmk-Adj.	54.12%	8.70								1.50	1.72	1.39
Excess Return	51.32%	0.77									-1.05	-0.34
Info. Ratio	53.07%	7.13										0.63
Sharpe Ratio	52.19%	1.56										

Table 19: Flow Sign Test - US Separate Account Composites, Active, 2012-2019

Flow sign test results for actively managed US separate account composites using only composites with strictly consecutive monthly data over the full sample period 1991-2019. The table shows a comparison between model sign and flow sign agreement. The first two columns report the average percent of times that model alpha sign predicts flow sign in the following period, and the statistical significance of this result. The remaining columns show the statistical significance of pairwise tests between the models. Standard errors are double clustered by fund and time.

Flow Sign Test - US Separate Account Composites - 2012-2019

	$\frac{\beta+1}{2}$	T-stat	Rating ≥ 4	Rating ≥ 3	CAPM	Mkt-Adj	FF3	FF4	Bmk-Adj	Exc Ret	IR	SR
Rating ≥ 5	61.94%	15.78	6.84	7.04	5.51	8.78	5.99	5.86	9.82	3.49	11.36	9.04
Rating ≥ 4	57.34%	14.8		3.10	2.89	5.63	3.45	3.53	7.46	2.06	9.02	6.55
Rating ≥ 3	56.02%	13.17			0.49	3.48	0.98	1.14	4.59	1.6	6.5	5.14
CAPM	53.95%	10.27				2.24	-0.47	-0.14	1.13	3.38	2.78	8.46
Mkt-Adj.	53.80%	12.37					-2.31	-1.89	-0.04	1.88	3.60	5.71
FF 3-factor	54.08%	11.34						0.66	2.05	3.46	3.11	8.21
FF 4-factor	53.99%	11.05							1.64	3.33	2.90	8.10
Bmk-Adj.	53.80%	12.04								1.88	3.63	5.40
Excess Return	51.76%	1.83									-0.34	2.65
Info. Ratio	51.95%	4.70										2.77
Sharpe Ratio	49.99%	-0.02										

Appendix B. Database Cleanup and Merge

In this data appendix, I detail the process of loading, merging, and cleaning all of the datasets used in this paper.

Returns data

I start with the monthly gross returns dataset for Separate Accounts. To obtain this dataset, I first pull monthly gross returns from the United States Separate Accounts database in Morningstar over the period 01/01/1980 - 09/30/2020. I do not limit the dataset to only surviving instruments, but rather include any separate account composite that existed at any time over the period 1980-2020. I require the composite to have continuous source data. The initial dataset contains 17,793 composites. I remove any composite that doesn't have a single return data point over its entire history, after which I have 16,518 composites. After removing a further 7 composites whose only existing return datapoints are 0's, there are 16,511 composites remaining.

Next, I pivot the dataset so that each observations is of the form (SecId, Date, Gross Return). Note that SecId is the unique Morningstar identifier for a share class of an investment. All share classes of the same fund are linked together using the FundId. Some funds do not have multiple share classes, which is the case for all Separate Account Composites. I drop any observations with missing return, and end this step with 2,388,151 observations.

I load, clean and pivot the monthly net returns dataset from the United States Separate Accounts database in Morningstar over the period 01/01/1980 - 09/30/2020 analogously. After the loading, cleaning and pivoting steps I end with 2,348,386 observations.

I proceed similarly for the monthly gross returns dataset for Mutual Funds. First, for the monthly gross return dataset I exclude share classes without a single month of return data. Next, I pivot the return dataset to be of the form where each observation is a (SecId, Date, Gross Return) triple. I drop any such observations that have missing gross return values. This leaves me with 6,340,742 month - share class observations at this stage. Note that at this point I am including the entire 01/01/1980 - 09/30/2020 time period.

Assets data

In this section, I describe the process of loading, cleaning, and pivoting the asset datasets.

For the Separate Accounts asset dataset, I first load the daily Net Assets data for United States Separate Accounts from Morningstar over the period 01/01/1980 - 09/30/2020. This dataset includes 17,079 separate account composites. Next, I remove any composite that doesn't have a single month of asset data, leaving 15,633 composites.

I drop any dates that do not have holdings for any fund (this helps to keep the size of the daily assets dataset manageable), and then pivot the dataset to so that each observation is of the form (SecId, Date, Net Assets). At this point, the dataset contains 16,289,586 such observations. Next, I drop any observations where net assets are missing or equal to 0. This leaves me with 1,170,212 observations at this stage. Limiting the dataset to observations whose dates are month-end dates, I am left with 1,169,154 observations.

I next limit my data sample to the time period 01/01/1991 - 09/30/2020, which leaves 1,155,001 observations for 15,501 composites. I also exclude, following Barber et al. (2016) and Ben-David et al. (2020), any observations where net assets are less than $10 million. I am then left with 929,824 observations for 12,326 different composites.

Merged dataset

In the first step of the merging process, I merge the return and assets datasets described above. I allow only observations that have a gross monthly return value, so the initial merged dataset has 2,388,151 observations, same as the cleaned return dataset. I then drop any observations for whose composite there is not a single assets data point over its entire history. After this step, I am left with 2,012,929 observations for 12,138 composites.

I then proceed to merge my returns and assets dataset with net expense ratios, turnover ratios, and dividends. For each data item, I first acquire the dataset; then pivot into (date, share class, data item) format; and finally, merge it with the existing dataset on (date, share class) values. When data is available on an annual basis, such as for net expense ratios and turnover ratios, I set the same data value for all months in a given year. With these merges, in case a data point is missing for a particular month - share class, I set the data point value to missing. I have 544,099 year - share class net expense ratio observations, 806,338 year - share class turnover ratio observations, and 2,721,396 month - share class dividend observations. Note that in this way, after the end of the merges, I am still left with 1,485,039 month - share class observations as before.

I then merge in the share class characteristics file, which includes data points regarding oldest share class, base currency, Morningstar category, and so forth. Once again, I keep every one of the 1,485,039 observations during the merge; if any characteristic data point is missing for a given month - share class pair, I set its value to missing.

Next, I merge the Morningstar ratings data with the existing dataset of returns, assets, expenses, dividends, turnover ratios and characteristics. In the merge I only include data points that have returns and assets, but might be missing Morningstar ratings. After the merge, as before I am left with 510,863 month - share class observations.

A final step of my data cleaning procedure is to consider only US Equity funds. I ensure I am looking only at US Equity funds by pulling data from the US Open End Fund database in Morningstar, and further filtering out funds based on the following conditions: (i) Global Broad Category Group is 'Equity'; (ii) US Category Group is 'U.S. Equity'; and (iii) Base Currency is 'US Dollar'. Filtering out by Global Broad Category Group yields 783,054 remaining month - share class observations. After filtering out by US Category Group, I am left with a further 537,485. Finally, after applying the Base Currency filter I have 537,485 remaining month - share class observations. In a final cleaning step for this portion of the data setup, I exclude any share classes benchmarked to a non-MSCI, SP, or Russell benchmark. The reason for this is that

US Mutual Funds are predominantly benchmarked to indices of these three providers, and I want to exclude very specific custom benchmarked-funds from this analysis in order to avoid unforeseen biases. After filtering the MSCI, SP, and Russell benchmarked share classes, I am left with 510,863 month - share class observations for the mutual fund dataset and 711,221 month - composite observations for the separate accounts dataset.