How Do Bond Investors Measure Performance? Evidence from Mutual Fund Flows*

Thuy Duong Dang[†], Fabian Hollstein[†], and Marcel Prokopczuk[†,‡]

May 16, 2019

Abstract

Which factor model do investors in corporate bonds use? We examine this question by tracking investors' decisions to invest in actively managed corporate bond mutual funds with a revealed preference approach. Our main result is that *all* bond factor models are dominated by the simple Sharpe ratio. For all corporate bond mutual fund styles, the Sharpe ratio explains fund flows better than alphas from bond factor models. Since the Sharpe ratio can be manipulated, our findings have potentially severe implications for both corporate bond mutual fund managers and investors.

JEL classification: G11, G12, G23

Keywords: Bond factor models, Sharpe ratio, bond mutual funds, investor flows, performance evaluation, flow-performance sensitivity

---

*Contact: dang@fmt.uni-hannover.de (T.D. Dang), hollstein@fmt.uni-hannover.de (F. Hollstein) and prokopczuk@fmt.uni-hannover.de (M. Prokopczuk). Part of this research project was finished when Marcel Prokopczuk visited Cass Business School, the University of Technology Sydney and Macquarie University, Sydney.

[†]School of Economics and Management, Leibniz University Hannover, Koenigsworther Platz 1, 30167 Hannover, Germany.

[‡]ICMA Centre, Henley Business School, University of Reading, Reading, RG6 6BA, U.K.

I Introduction

Corporate bonds are an important, yet underresearched asset class. While the total market scale is somewhat lower than that of equities, the annual issuance of corporate bonds is at a remarkably larger scale (by both value and number of issues) than that of equity for US corporations: for example, in 2017, there were 3,329 corporate bond issues totaling $2.5 trillion compared to 1,069 stock issues totaling $223.5 billion.[1]

In this paper, we adopt a revealed preference approach as in Barber et al. (2016) and Berk and Van Binsbergen (2016) to address the following research questions: How do investors evaluate corporate bonds? Which factor model do they use to measure the performance of corporate bond mutual funds? Do investors even use factor models or do they rely on simpler performance measures?

We employ mutual funds as an instrument to investigate which factor models investors use in corporate bond markets. Through most of the past decade with several periods of market turmoil, bond mutual funds have experienced net inflows, while equity funds have continuously experienced net outflows (see Figure 1).[2] Bond mutual funds thus have become an important investment vehicle, in particular for individual investors, that seek exposure to bond markets.[3]

We begin our empirical analysis by conducting a flow-performance horse race to infer which performance measures corporate bond investors use when allocating capital. At one extreme, investors may simply rank funds based on their raw returns; at the other extreme, they may rank funds based on the alpha from a multi-factor model for

[1]SIFRMA Fact Book 2018, sources: Bloomberg, Dealogic, Thomson Reuters. Available at https://www.sifma.org/wp-content/uploads/2017/08/US-Fact-Book-2018-SIFMA.pdf

[2]See Investment Company Fact Book (2018). Available at https://www.ici.org/pdf/2018_factbook.pdf

[3]One of the prominent trends in recent years is the fast growth of assets under management by fixed income mutual funds. According to the Investment Company Institute Fact Book, total net assets of corporate bond mutual funds increased dramatically from $355.63 billion in 2000 to $2.21 trillion in 2017. Demographics influence the demand for bond mutual funds and have supported bond fund flows over the past decade. Older investors are likely to have higher account balances because they have had more time to accumulate savings and tend to shift toward fixed-income products. Since the 2008-2009 financial crisis, many investors have shifted their investment from equity funds to bond funds, which reflects investors' perceptions about financial conditions.

returns. Given the significant model uncertainty associated with evaluating corporate bonds, we measure performance using a range of single and multi-factor models as well as ratios that are commonly found in the academic literature on asset pricing. For our main analysis, we use the best-known and most widely used measures/ models, including: the raw return, the Sharpe ratio, a single factor model with a bond market index, a two-factor model with a stock and a bond market index, the Elton et al. (1995) four-index model and the Fama and French (1993) five-factor model for bonds. We test which of the performance measures best explains investor flows, while controlling for well-known predictor variables such as lagged fund flows, expense ratios, fund size, age, and Morningstar ratings.

Our main contribution is thus a systematic analysis of which factors investors in corporate bond markets use to assess performance. To the best of our knowledge, we are the first to comprehensively analyze this and the questions stated above.

We find that the Sharpe ratio consistently explains corporate bond fund flows significantly better than the raw return or any alpha from a single- or multi-factor model. This is not only true for the full sample, but also for the subsamples of high yield and investment grade funds.

Next, we test whether more sophisticated investors use more sophisticated benchmarks or models to assess mutual fund performance when making their investment decisions in corporate bonds. As expected, the Sharpe ratio can explain fund flows of retail-oriented funds best. It is aligned with the result of Chakraborty et al. (2018) that mutual fund retail investors face information constraints and thus utilize salient and plausibly relevant available information. More surprisingly, even for institutional-oriented corporate bond mutual funds, the Sharpe ratio also explains fund flows better than any factor model, although not significantly so relative to every factor model.

In a further step, we investigate whether investors seek exposure to systematic risk factors. Our findings are consistent with the evidence documented by Gebhardt et al. (2005a) that default and term risk are important determinants of corporate bond returns and bond investors' behavior documented by Chen and Qin (2016) that investor flows to corporate bond funds are sensitive to macroeconomic conditions. Investors tend to divest their capital when default risk is high and corporate bond funds attract flows when term spreads are high.

We run a battery of robustness tests. Our results are qualitatively similar for (i) a longer horizon for performance evaluation (one year); (ii) different ways to calculate the Sharpe ratio; (iii) the Berk and Van Binsbergen (2016) testing approach; (iv) using various alternative factor models: for example, the Ludvigson and Ng (2009) bond macro factor model, as well as the bond factor model recently suggested by Bai et al. (2019); (v) the inclusion of the Morningstar fixed income style box as additional control variable.

Wermers (2011) reviews several important advances in the measurement of performance of actively managed portfolios. The return-based model most widely used among academics in analyzing equity managers is the four-factor model of Carhart (1997). For hedge funds, the seven-factor model of Fung and Hsieh (2004) appears to be favored by academics. On the other hand, there is no consensus about risk factors and factor models for fixed-income markets.

The earlier research on corporate bond returns generally rely on long-established stock and bond market factors, including the stock market factors of Fama and French, (1993; 2015), Carhart (1997), and Pástor and Stambaugh (2003); accompanied by the bond market factors of Fama and French (1993), Elton et al. (1995). However, these commonly used factors are either constructed from stock-level data or aggregated indices and macroeconomic variables. Recent studies of, for instance, Jostova et al. (2013) on bond momentum; Lin et al. (2011) on bond liquidity risk; Chung et al. (2019) on volatility, and Bai et al. (2019) on bond downside risk, credit risk and liquidity risk rely on the features of corporate bonds when constructing bond risk factors to explain the cross-sectional differences in corporate bond returns.[4]

All studies mentioned above suggest different models for corporate bonds. The authors mainly examine how well these models perform in explaining the cross-section of corporate bond returns. We contribute to this literature by examining which of the many factor models proposed in the previous literature are actually used by investors.

We also contribute to the bond fund literature. Results from studies of Zhao (2005), Comer and Rodriguez (2013) and Fulkerson et al. (2013) suggest that bond fund investors respond to risk-adjusted measures rather than the raw return. However, those previous studies do not examine *which* risk-adjusted measure is used by the majority of investors in corporate bond funds. We complement these studies by examining which performance measure actually drives the flow sensitivity.

We further extend the research on equity mutual funds (Barber et al., 2016 Berk and Van Binsbergen, 2016 and Ben-David et al., 2019) and hedge funds (Agarwal et al., 2018 and Blocher and Molyboga, 2017) to corporate bond mutual funds to offer an integrated view of investor behavior in the mutual fund industry. The recent evidence for both equity mutual funds and hedge funds documents that the CAPM alpha is more successful in explaining fund flows over alphas of other multi-factor models. Corporate bond fund investors likely act differently from equity fund investors because corporate bonds and stocks have different return and risk characteristics. In addition, the model uncertainty is substantially greater for corporate bond investors: the bond literature has not yet settled even on a class of main factor models, not to mention details about which and how many risk factors to use.

Our results carry potentially important implications for investors and fund managers. Unlike equity funds, corporate bond funds tend to hold more illiquid assets whose prices are stale. Smoothing returns over time will remain the portfolio's mean

[4] Other research on bond factors includes Houweling and Van Zundert (2017) on size, low risk, value and momentum factors of corporate bonds; Israel et al. (2018) on carry, defensive (low risk), momentum and value; Bektic et al. (2017) on size, value, profitability and investment factors for corporate bonds.

return unchanged but decrease its variance, which biases the Sharpe ratio and other similar performance measures upward (Bollen and Pool, 2008 and 2009; Getmansky et al., 2004). If investors tend to prefer using the Sharpe ratio as the main performance measure to evaluate funds, then fund managers who are aware of this will have an obvious incentive to take actions that enhance these measures without adding real economic value. Funds with illiquid assets, whose prices are only reported occasionally may benefit from this. Reported performance could therefore strongly mislead investors' decisions. Fund managers who want to manipulate their Sharpe ratios can do so by holding more illiquid assets or "mis-marking" bonds, which creates trading opportunities for active traders of mutual funds and in return, poses the threat to both those managers and buy-and-hold investors of the funds.

The remainder of this paper is organized as follows. Section II describes our data and the estimation methods. Section III conducts the flow-performance horse race. In Section IV, we show further evidence and test the robustness of our main findings. Section V discusses the implications of our results. Section VI provides concluding remarks and suggestions for further research.

II Data and Method

A Data

Our data on U.S. actively managed corporate bond funds comes from the Center for Research in Security Prices (CRSP) Survivor-Bias-Free U.S. Mutual Fund Database. We use data from 1991 to 2017.[5] Since we use an estimation window of five years in our empirical analysis, our final sample period used for testing is from 1996 to 2017.

A bond fund often offers several share classes with different combinations of expense ratios, management fees, front-end and/or back-end sales charges (load), minimum

[5]We choose 1991 as starting data because monthly information on fund size became available at that time. Furthermore, prior to 1991, there are only few corporate funds in the database.

investment requirements, as well as restrictions on investor types. These different share classes are designed to attract investors with different wealth levels and investment horizons. Since these fund-share level characteristics can influence the investment and redemption decisions of mutual fund investors, we follow Goldstein et al. (2017), Chen et al. (2010a) and Jiang and Yuksel (2017) and use individual fund share classes as our unit of observation.

We select corporate bond funds based on the objective codes provided by CRSP.[6] Because we are interested in investors who are attempting to identify managerial skill in their fund allocation decisions, we exclude index funds, exchange traded funds and exchange traded notes. We remove funds with TNA less than $10 million and age of less than three years to mitigate data biases, such as incubation bias (Evans, 2010). Our final sample includes 1,482 unique funds and 3,647 unique share classes (1060 high yield share classes and 2587 investment grade share classes). We merge the CRSP data with the Morningstar Direct database, matching on fund CUSIPs and Tickers following Pástor et al. (2015) and Berk and Van Binsbergen (2015).

To measure the performance of corporate bond funds, for our main tests, we use the raw return, the Sharpe ratio and five different models that investors might reasonably employ for performance evaluation. The Sharpe ratio is probably the best-known and most widely used measure of portfolio performance employed in the fund industry.[7] The factor models used include a single factor model of the aggregate bond market return (CAPM bond), a 2-factor model of both the aggregate bond and stock market return (CAPMstkb) following Goldstein et al. (2017), the Elton et al. (1995) 4-factor model (E4) including the aggregate bond (MKT^{bond}) and stock (MKT^{stock}) market

[6]Specifically, to be classified as a corporate bond fund, a mutual fund must have a (i) Lipper object code in the set ('A', 'BBB', 'HY', 'SII', 'SID', 'IID') or (ii) Strategic Insight objective code in the set ('CGN', 'CHQ', 'CHY', 'CIM', 'CMQ', 'CPR', 'CSM'), or (iii) Wiesenberger objective code in the set ('CBD', 'CHY') or (iv) 'IC' as the first two characters of the CRSP objective code.

[7]See Goetzmann et al. (2007) and Elton and Gruber (2013). The Sharpe ratio is used, for example, in the Schwab Select List and the Standard and Poor's Select Funds mutual fund rating and in the *Hulbert Financial Digest* newsletter ratings.

excess returns, default risk (DEF) and option (OPTION) factors, the Fama and French (1993) 5-factor model (FF5) including 3 common stock factors: MKT^{stock}, the size factor (SMB),the value factor (HML) and 2 bond factors: the term spread (TERM) and DEF. Elton et al. (2001) show that in addition to expected default, state taxes, the Fama and French stock factors explain the rate spread on corporate bonds. Empirical evidence in Gebhardt et al. (2005a) and Lin et al. (2011) show that DEF and TERM betas are important determinants of required corporate bond returns. Finally, we consider an augmented FF5 model that adds the liquidity (LIQ) and momentum (MOM) factors (FF7).[8] In many cases, these models yield similar rankings of mutual funds (i.e., the performance measures are highly correlated). However, we exploit the cases in which rankings differ across models to answer the question which performance measure best explains the choices that investors make when allocating capital to actively managed corporate bond mutual funds. We focus on return-based approaches (rather than portfolio holding-based approaches) to measure performance because they rely on less information from fund managers and return data are usually available on a much more frequent basis in case of mutual funds.

The excess bond market return (MKT^{bond}) is proxied by Barclays US Aggregate Bond Index in excess of one-month T-bill return. TERM is defined as the difference between the monthly long-term government bond return and the one-month Treasury bill rate, which captures returns generated by increasing duration (i.e., higher interest rate risk). DEF is defined as the difference between the return on a high yield bond index and intermediate government bond return, capturing returns generated by taking on higher default risk. OPTION captures nonlinearities due to investment in mortgage backed securities and is measured by the difference between Barclays GNMA index and

[8]The factors MKT^{stock} (excess market return), SMB (small minus big), HML (high minus low), MOM (winners minus losers), LIQ (liquidity risk) are described in and obtained from Kenneth French's and Lubos Pastor's online data libraries: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html` and `https://faculty.chicagobooth.edu/lubos.pastor/research/liq_data_1962_2017.txt`

Barclays Government Intermediate Index. We obtain monthly return data of Barclays bond indices from Morningstar Direct.

We use equity MOM and LIQ factors in our main test instead of the Jostova et al. (2013) bond MOM and the Lin et al. (2011) LIQ factors for two reasons: First, the factors are only available from the authors for a subset of our sample period. The equity factors are available for our entire sample period.[9] Second, Lin et al. (2011)'s results show that the coefficient of the Pástor and Stambaugh (2003) stock liquidity factor beta is significant even after incorporating bond characteristic variables, suggesting a possible cross-market liquidity risk effect. Jostova et al. (2013) and Gebhardt et al. (2005b) find significant evidence of a momentum spillover from equities to corporate bonds (i.e., past equity returns significantly predict future bond returns). However, we also conduct robustness tests using bond LIQ and MOM factors as well as the Bai et al. (2019) factor model for a shorter sample period corresponding to the period for which those factors are available in Section IV.

B Empirical Approach

1 Fund Flows

The key variables in our empirical analysis are mutual fund flows and the different performance measures. Following standard practice, we calculate flows of fund p in month t as the percentage growth of new assets, assuming that all flows take place at the end of the month:

$$F_{pt} = \frac{TNA_{p,t}}{TNA_{p,t-1}} - (1 + R_{p,t}), \tag{1}$$

[9] Bond MOM data is provided from 1974 until June 2011 on Gergana Jostova's website `https://business.gwu.edu/gergana-jostova` Bai et al. (2019) propose a new bond factor model which includes: downside risk (DRF), credit risk (CRF) and liquidity risk (LRF). Data on Corporate Bond Risk Factors are available on Turan Bali's website: `http://faculty.msb.edu/tgb27/workingpapers.html`. DRF and CRF cover the period from July 2004 to December 2016, LRF from August 2002 to December 2016. Using the TRACE database to create the factors ourselves would also limit the sample period, because TRACE does not start before July 2002.

where $TNA_{p,t}$ is the total net assets under management of fund p at the end of month t, and $R_{p,t}$ is the return of fund p in month t. To mitigate the influence of outliers, we follow Goldstein et al. (2017) and winsorize fund flows at the 1% and 99% levels.

2 Fund Performance Measures

We proceed in two steps to estimate the realized alphas. First, we estimate the abnormal return (alpha) for each mutual fund using each of models. Alpha estimates are updated monthly based on a rolling estimation window. In the following, we will outline the procedure in more detail for the FF5 model. The procedure is similar for all other factor models. We obtain factor loadings by the following time-series regression using 60 months of return data for months $\tau = t-1$ until $t-60$:

$$(R_{p,\tau}-R_{f,\tau}) = \alpha_{pt}+\beta_{pt}MKT_{\tau}^{stock}+s_{pt}SMB_{\tau}+h_{pt}HML_{\tau}+t_{pt}TERM_{\tau}+d_{pt}DEF_{\tau}+e_{p\tau}. \quad (2)$$

The parameters β_{pt}, s_{pt}, h_{pt}, t_{pt}, and d_{pt} represent the exposure to stock market, size, value, term risk and default risk, respectively, of fund p at time t; α_{pt} is the mean return unrelated to factor tilts, and $e_{p\tau}$ is a mean zero error term.

We then calculate the alpha for the fund in month t as its realized return less the model-implied return in month t:

$$\hat{\alpha}_{pt} = (R_{pt}-R_{ft}) - \left[\hat{\beta}_{pt}MKT_{t}^{stock} + \hat{s}_{pt}SMB_{t} + \hat{h}_{pt}HML_{t} + \hat{t}_{pt}TERM_{t} + \hat{d}_{pt}DEF_{t}\right], \quad (3)$$

For other factor models, we adjust Equations (2) and (3) accordingly.

The Sharpe ratio of fund p at the end of month t is calculated as the ratio of the excess return of fund p at the end of month t over the standard deviation of its monthly returns over the past year. In Section IV, we also test the robustness of our main results to alternative calculations of the Sharpe ratio.

3 Horizon for Performance Evaluation

With rational expectations, investors respond to their perceptions about the skill of a fund manager. With new information, they should update this perception. However, how investor should weight past returns when assessing fund manager skill is less clear. To make a decision about what performance horizon to analyze when comparing models, first, we estimate the following simple model of the flow-return relation:

$$F_{pt} = a + \sum_{s=1}^{S} b_s R_{p,t-s} + e_{pt}, \tag{4}$$

where F_{pt} are fund flows for fund p in month t and $R_{p,t-s}$ represents the lagged returns for the fund at lag s, where we vary the number of maximum lagged return from $S = 1$ to 48 months. The Akaike information criterion (AIC) yields a minimum for S=1.[10] We thus settle on a lag length of 1 month. In Figure 2, the black line depicts the estimated b_s coefficients (y axis) at a various lags (x axis). From this figure, it becomes clear that the first lag return is the most influential indicator about fund performance to investors, while the sensitivities to more distant returns are close to zero.

C Descriptive Statistics

In Table 1, we provide summary statistics. In total, our sample consists of 352,243 share class-month observations for 3,647 unique share classes from January 1996 to June 2017. Panel A of Table 1 summarizes the fund characteristics. The average fund share class has total net assets of about \$609.15 million, though the median is considerably smaller with \$113 million, which suggests that fund size is skewed by large funds. The mean (median) fund age is 10.61 (8.59) years. The average annual expense ratio for our sample is 0.96%. A large proportion of funds (75%) has either a front-end or back-end load. The average fund return standard deviation amounts to 1.22%, which is substantially lower compared to that of equity funds (Barber et al., 2016 report 4.92% for their sample).

Panel B of Table 1 reports descriptive statistics of monthly fund flows and returns, the two

[10]The Bayesian information criterion (BIC) also yields a minimum for S=1. Both AIC and BIC also yield a minimum for S=1 when we additionally include control variables and time fixed effects.

key variables of our analysis. Over the sample period, the mean return (the time-series average of the cross-sectional distribution of monthly fund returns) of all share classes in our fund sample is 0.42% per month (5.04% per annum). Investment grade bond funds yield an average return of 0.37%, while high yield bond funds yield an average return of 0.55% per month. The average (median) of the percentage fund flow is 0.58 (–0.11), with a standard deviation of 5.09% per month. The dispersion in fund flows is higher than that documented for equity funds, 2.25% by Barber et al. (2016). High yield bond returns exhibit an average first order correlation of 23.07%, which is higher than that of investment grade funds (16.15%).[11] There is also a high level of serial correlation in the flow ratio. The first-order autocorrelation is 28%, which is approximately equal for high-yield and investment-grade funds. As can be seen from Panel D of Table 1, the correlations between Morningstar rating and other performance measures are quite low, under 10%.

III Empirical Results

A Model Horse Race

As in Barber et al. (2016), we classify fund performance into decile ranks and examine in a pairwise fashion which model better explains the flows when the models yield different performance ranks. We estimate the relation between fund flows and a fund's decile ranking based on two different performance measures by estimating the following regression:

$$F_{pt} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{pt-1} + \mu_t + e_{pt}, \quad (5)$$

where the dependent variable F_{pt} is the fund flow of mutual fund p in month t. In each month of our test period, we assign the decile performance rank for each fund based on each of the measures.[12] Decile 10 includes the best performing funds and Decile 1 contains the

[11] Analyzing daily returns, Chalmers et al. (2001) find that the average bond autocorrelation is higher than that of equity funds and the equity index.

[12] We rank a fund's performance based on each performance measure within the fund's category peer group (i.e. investment grade or high yield). This ensures that the rankings are driven mainly by managerial skill rather than choice of investment style or systematic events that affect all funds in a category peer group.

worst funds based on the performance measure. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund p in month $t-1$ is in the decile i based on the first model and decile j based on the second model. To estimate the model in Equation (5), the dummy variable for $i = 5$ and $j = 5$ is excluded. The matrix $X_{p,t-1}$ represents control variables including the lagged fund flow from month $t-1$, the lagged fund expense ratio, a dummy for no-load funds, a fund's return standard deviation estimated over the prior twelve months, the log of fund size and the log of fund age in month $t-1$ as well as Morningstar rating dummies in month $t-1$.[13] We also include time fixed effects (μ_t). Following Petersen (2009) and Cameron et al. (2011), we double-cluster the standard errors by fund and month. Clustering by funds helps address serial correlation in residuals and, more importantly, over time among the different share classes of a given fund. Clustering by month helps address cross-sectional correlation in residuals across different share classes at a given time point.

The key coefficients of interest are b_{ij} ($i = 1, 2, ..., 10$ and $j = 1, 2, ..., 10$), which can be interpreted as the percentage flows received by a fund, which is in decile i based on the first performance measure and in decile j for the second measure relative to a bond fund that ranks in the fifth decile for both performance measures. With each pair of coefficients b_{ij} and b_{ji} to determine whether investors are more sensitive to the first or to the second measure (alpha estimated using the first model or using the second model), we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. If investors are more responsive to the first measure than the second one, we would expect to reject the null hypothesis in favor of the alternative hypothesis that b_{ij} is greater than b_{ji}. For each pairwise comparison of two performance measures, we have 45 such b coefficient comparisons. We use a Wald test to test the null hypothesis that the summed difference across all 45 comparisons is equal to zero, and calculate a binomial test statistic to test the null hypothesis that the proportion of differences equals 50%. We present a model as "winning model" if the sum of coefficient differences is statistically significant from zero.

[13] Note that we expand the set of control variables in Barber et al. (2016) because Del Guercio and Tkac (2008), Ben-David et al. (2019) and Evans and Sun (2018) show that that Morningstar ratings substantially influence the allocation decisions, in particular for retail investors. Morningstar is the dominant information intermediary among financial advisors, being more influential than, for example Lipper and Standard & Poor's.

Table 2 reports the main results. In Panel A, we compare the Sharpe ratio to all other performance measures. We find that compared to the raw return, the sum of coefficient differences amounts to 17.42, which is highly statistically significant. 91.11% of the coefficient differences are positive. It is thus clear that investors are substantially more responsive to the Sharpe ratio than they are to the raw return. The pairwise comparisons of the Sharpe ratio with all factor models yield similar results. The sum of coefficient differences with the CAPMb amounts to 14.16 and that with CAPMstkb to 9.85, which are both also highly statistically significant. The differences are comparably smallest between the Sharpe ratio and the E4 model, where the sum of coefficient differences amounts to 6.54. However, this is also highly statistically significant and 73.33% of the coefficient differences are positive. For the FF5 and FF7 models, the results are even more pronounced and highly statistically significant. Our first main result thus strongly suggests that investors rely on the Sharpe ratio rather than any more sophisticated factor model when making their capital allocation decisions.

Considering factor models, we find generally greater flows to bond mutual funds with higher ranks based on the Elton et al. (1995) four-factor model alpha than to that based on the other competing models. Table 2 indicates that around 70%-80% of the cases, investors allocate more capital to corporate bond mutual funds when the E4 alpha performance rank exceeds the alphas estimated using competing factor models. The E4 model explains investor fund flows significantly better than the raw return and any other factor model. Given the outperformance of the simple Sharpe ratio over the model, this results, though, appears to be of minor importance.

Our horse race tests are similar in spirit to recent studies by Barber et al. (2016) and Berk and Van Binsbergen (2016). It is natural to compare our corporate bond mutual fund evidence with the findings documented for equity mutual funds and hedge funds. Both Barber et al. (2016) and Berk and Van Binsbergen (2016) for equity mutual funds, and Agarwal et al. (2018) and Blocher and Molyboga (2017) for hedge funds, show that the CAPM alpha explains investor flows better than the raw return or alphas from any factor model. Our results for corporate bond mutual funds contrast these. Both CAPM-style models (a single factor bond market model and a two-factor stock–bond CAPM) cannot beat other performance measures. Even more strikingly, the Sharpe ratio even beats all factor models for corporate bonds.[14]

B Tests on Subsamples of Corporate Bond Funds

Our primary analysis treats all corporate bond mutual fund investors as a homogeneous group. However, different groups of investors may use different methods to assess the performance of funds. In this section, we test whether there are differences between investor flows to (i) high yield and investment grade funds and (ii) retail and institutional-oriented funds.

1 High Yield and Investment Grade Bond Funds

We test whether investors in investment grade and high yield bond funds differ in the way they evaluate performance of the fund when making capital allocation decisions. To perform this analysis, we split the entire sample into separate high yield and investment grade groups. Following Chen and Qin (2016), we categorize funds with Lipper object code "HY", Strategic Insight objective code "CHY" or Wiesenberger objective code "CHY" as high-yield bond funds. Funds with all other objective codes are classified as investment-grade bond funds.

Tables 3 reports the results for high yield bond funds, whereas Table 4 presents the results for investment grade funds. One can observe that the Sharpe ratio is able to best explain variation in flows across both high yield and investment grade bond mutual funds. Thus, investors in both classes appear to use the Sharpe ratio as primary performance measure which they base their capital allocation decisions on. Regarding factor models, investors of high yield bond funds are more sensitive to alphas of models that include the default risk factor than factor models which do not include this factor whereas investors of investment grade bond funds seem to seem to be more sensitive to the abnormal return relative to the bond (and stock) CAPM than to that of more complex models.

[14]Note that Barber et al. (2016) do not present the result of pairwise comparisons between the Sharpe ratio and other measures. Agarwal et al. (2018) and Berk and Van Binsbergen (2016) do not consider the Sharpe ratio at all in their empirical studies.

2 Retail and Institutional-Oriented Bond Funds

Corporate bond market markets are dominated by institutional investors, such as insurance companies, pension funds and mutual funds. On the other hand, according to Flow of Fund data during the 1986-2012 period, about 82% corporate bonds were held by institutional investors. In contrast, interestingly, according to the ICI fact book, total net assets of corporate bond mutual funds held by institutional accounts are only around one-tenth of those held by primary accounts of individuals issued by a broker-dealer.

We split the sample into retail investor and institutional investor-oriented funds by the classification provided by CRSP Mutual Fund Database. From December 1999, CRSP assigns each fund share a dummy for institutional share and a dummy for retail share.[15] The main criteria used to classify are the minimum investment requirement and the distribution channel.[16] The two dummies are not mutually exclusive. Therefore, we set a fund share as institutional-oriented if the CRSP institutional share dummy is one and the CRSP retail share dummy is zero.

Our results, presented in Tables 5, show that retail investors in bond mutual funds are most responsive to the Sharpe ratio among all performance measures. For retail-oriented mutual funds, the Sharpe ratio explains investor flows strongly and highly significantly better than any other performance measure. In Table 6, for institutional-oriented mutual funds, the Sharpe ratio also explains investor flows better than any other model. While the sum of coefficient differences is positive in any case, however, the differences are only statistically significant when compared to the return, alphas of the CAPM bond, FF5, FF7 models. Moreover, for institutional-oriented mutual funds, all factor models explain investor flows significantly better than the simple raw return.

[15] We backfill the CRSP investor-oriented fund classification for those funds.

[16] According to the ICI Fact Book, institutional accounts include accounts direct-sold, purchased by an institution, such as business, financial organizations. Accounts of individuals are issued by a broker-dealer. Morningstar classifies as institutional fund shares those typically purchased by large institutional buyers such as pension funds. These share classes are only offered to investors who invest $1 million or more with the lowest expenses in the mutual fund universe.

C Response of Investor Flows to Components of Fund Returns

The preceding analysis indicates that the Sharpe ratio best explains fund-flow relations overall and also every market segment considered. Huij and Derwall (2008) find strong evidence of performance persistence ("hot hand" effect) in bond funds (i.e., past performance predicts future performance). Funds that report strong (weak) performance in the past repeat their performance in the subsequent period. Jordan and Riley (2015) study equity mutual funds and show that past fund return volatility is a strong predictor of future fund performance. Low volatility funds have better performance than high volatility funds. Thus, investors in aggregate tend to consider the trade-off between risk and return of funds when evaluating fund performance.

However, this does not imply that investors completely ignore other factors that affect fund performance (e.g., term risk, default risk). In the previous section, we conduct horse race tests on subsamples and also find evidence that investors in high yield bond funds are more sensitive to alphas of models that include the default risk factor. In this section, we therefore examine whether investors consider factor-related returns when evaluating fund performance. We decompose each fund's excess return into its alpha and factor-related returns by rearranging Equation (3). We conduct the return decomposition analysis for two models: the Elton et al. (1995) four-factor model and a seven-factor model, which is an augmented Fama and French (1993) five-factor model added momentum and liquidity risk factors. For example, for the seven-factor model:

$$\begin{aligned}(R_{pt} - R_{ft}) \quad = \quad & \hat{\alpha}_{pt} + \Big[\hat{\beta}_{pt} MKT^{stock}{}_t + \hat{s}_{pt} SMB_t + \hat{h}_{pt} HML_t + \hat{t}_{pt} TERM_t \\ & + \hat{d}_{pt} DEF_t + \hat{m}_{pt} MOM_t + \hat{l}_{pt} LIQ_t\Big]. \end{aligned} \tag{6}$$

In this return decomposition, a fund's return is due to eight components: the fund's seven-factor alpha, the fund's exposure to stock market risk, size, value, term risk, default risk, momentum and liquidity risk. We calculate, for example, the portion of the fund's return related to term risk as: $TERMRET_{p,t-1} = \hat{t}_{p,t-1} TERM_{t-1}$. Using this return decomposition, we estimate the following panel regression across p funds and t months to test how investors react to different return components:

$$F_{pt} = b_0 + b_1 ALPHA_{p,t-1} + b_2 MKTRET^{stock}_{p,t-1} + b_3 SIZRET_{p,t-1} + b_4 VALRET_{p,t-1}$$
$$+ b_5 TERMRET_{p,t-1} + b_6 DEFRET_{p,t-1} + b_7 MOMRET_{p,t-1} + b_8 LIQRET_{p,t-1}$$
$$+ cX_{pt-1} + \mu_t + e_{pt}, \quad (7)$$

where the matrix of control variables, X_{pt-1} and month fixed effect μ_t are defined before. The parameter estimates of interest in Equation (7) are b_i, $i = 1, ..., 8$. For the Elton et al. (1995) model, we adjust Equations (6) and (7) accordingly. For investors who use a seven factor model (Elton et al. (1995) four-factor model) to estimate alpha to evaluate fund performance, we expect $b_1 > 0$. If investors value returns from fund exposure to any specific factor, we expect the b coefficient estimate corresponding to that factor to be significantly greater than 0.

The results are presented in Table 7. In Panel A, we base the return decomposition on the Elton et al. (1995) four-factor model and the results for the seven-factor model are in Panel B. We observe that the sensitivities of investor flows to alphas are significantly positive for both models and for all subsamples. For the high yield fund sample, the coefficient on the return component related to default risk is highly significantly negative, which implies that investors of high yield bond funds tend to shift their capital to funds with relatively less exposure to default risk. This is consistent with the finding in Chen and Qin (2016) that there is a positive association between the default spread and money flows into investment grade funds, suggesting that more capital flows into relatively safe funds when default risk is higher. It also may suggest that investors punish the risk-shifting behavior of fund managers (as in Huang et al., 2011) by taking higher risk when the default spread is higher. The other source of common risk for corporate bonds arises from unexpected changes in the term structure of interest rates. Flows to corporate bond funds show a significantly positive reaction to the return component of a fund's exposure to term risk, indicating that when the term spread is higher, corporate bond funds attract flows from investors. These results indicate that bond investors are aware of aggregate market conditions.

Surprisingly, flows to investment grade bond funds are positively sensitive to returns related to the value factor. Firms with high book-to-market ratio are usually large and more well-established companies with the underpriced stocks, hence may have good credit quality. Koijen et al. (2017) provide the evidence that returns on value stocks reflect compensation for macroeconomic risk and find that they are highly positively correlated with several bond market factors such as Cochrane and Piazzesi (2005) factor or the yield spread, which are leading indicators of business cycles. Furthermore, investors' flows to investment grade bond funds exhibits a strong preference for returns arising from funds' exposure to the equity momentum factor, which is consistent with the evidences documented in Avramov et al. (2007), Jostova et al. (2013) and Gebhardt et al. (2005b), that there is no momentum effect among investment grade bonds. Bond momentum is generated by high-yield bonds and there is a momentum spillover effect from equities to investment grade bonds (i.e., firms earning high (low) equity returns over the previous year earn high (low) bond returns the following year).

IV Robustness

A One-Year Horizon for Performance Evaluation

To test the robustness of our main results, in this section, we examine a one-year horizon, instead of one-month as in our main analysis. While our analysis based on the AIC indicates that a one-month window is optimal, one might argue that a one-year window is also suitable because it broadly balances relevance (i.e., recent returns are likely more informative) versus the signal-to noise ratio (i.e. short-term returns are mostly noise with very little signal about returns). Furthermore, there may be frictions such as inattention and transaction costs, which could create delays in the response of flows to fund performance.

The regression using Equation (4) yields a series of coefficient estimates, b_s, that represent the relation between flows in month t and the fund's return lagged s months, $s = 1, ..., 12$. Figure 2 shows that the most recent past return seems to be much more important to explain fund flows than more distant returns (i.e., the weights investors attach to past return quickly decay after the first previous month). Therefore, we follow Barber et al. (2016) and weigh the performance measures. We empirically estimate the rate of decay λ in the flow-return relation using an exponential decay model:

$$F_{pt} = a + b \sum_{s=1}^{12} e^{-\lambda(s-1)} R_{p,t-s} + e_{pt}. \tag{8}$$

We present the results of this regression in Figure 2. The smooth line presents the estimated decay function which closely tracks the unconstrained estimates from the regression of Equation (4). We apply this decay function to calculate each fund's alphas as weighted average of the prior twelve monthly alphas:

$$ALPHA_{p,t-1} = \left(\frac{\sum_{s=1}^{12} e^{-\lambda(s-1)} \hat{\alpha}_{t-s}}{\sum_{s=1}^{12} e^{-\lambda(s-1)}} \right), \tag{9}$$

where monthly alpha estimates are based on one of the five models that we evaluate and the exponential decay rate is based on the estimates from Equation (8).

The Sharpe ratio of fund p at the end of month t is calculated as the ratio of weighted average of prior 12-month excess return of fund p using the decay rate λ over its return's one-year standard deviation.

Table 8 reports the result of the test using the one-year performance horizon. Consistent with the one-month horizon, the Sharpe ratio explains fund flows significantly better than the raw return and any of the factor models.

B Alternative Measurement of the Sharpe Ratio

As an alternative to the simple volatility estimate, we employ a GARCH (1,1) model using 60-month past returns to estimate the denominator of a fund's Sharpe ratio to align the estimation period with the method to estimate alphas. We then repeat our model horse race. The results as reported, in Table 9, are qualitatively similar to those from our main test. Further untabulated analyses reveal that our main results are also robust when we consider the Sharpe ratio calculated as the average 12 monthly fund returns over its standard deviation.

C Alternative Factor Models

We check the robustness of our main results with a battery of alternative relevant recent factor models for bonds, including[17]:

1. An augmented Fama and French (1993) model with the Jostova et al. (2013) bond momentum factor (MOMb).

2. The Bai et al. (2019) four-factor model including a bond market factor and three new factors: downside risk, credit risk and liquidity risk (B4).

3. An augmented Fama and French (1993) model with liquidity risk and aggregate volatility risk as in Chung et al. (2019) (C7).

4. The Ludvigson and Ng (2009) macro-factors for bonds (macro).

5. An augmented Fama and French (2015) five-factor model with TERM and DEF (FF7e).

6. An augmented Hou et al. (2015) q-4 factor model with TERM and DEF (HXZ).

6. An augmented Stambaugh and Yuan (2016) M4 mispricing factor model with TERM and DEF (M4).

Table 10 reports the result of horse races between each alternative factor model and other measures used in our main test. Consistent with our previous results, the Sharpe ratio also explains investor flows significantly better than each of these additional factor models.

D Controlling for Morningstar Fixed Income Style Box

We also follow Aragon et al. (2019) and control for fund style effects by including Morningstar fixed income style box dummies as further control variables. The results reported in Table 11 clearly show that flows of investors into all different bond fund styles are most responsive to funds' Sharpe ratios.

[17]We test with the shorter sample length corresponding to the availability of data on those factors.

E The Berk and Van Binsbergen (2016) Testing Approach

For a final robustness test, we also employ the Berk and Van Binsbergen (2016) approach to examine the relation between flows and lagged fund performance. Instead of exploiting the full variation in fund flows, Berk and Van Binsbergen (2016) rely only on the signs of flows and performance measures.

First, we test for a positive relation between fund flows and performance (i.e., whether the regression coefficient of the sign of the subsequent flows on the sign of the performance measure is positive). Φ is defined as a simple sign function that returns the sign of a real number, taking values of 1 for a positive number, -1 for a negative number and 0 for zero. We test the following null hypothesis:

$$\beta_{flow,performance} = \frac{cov(\Phi(flow_{i,t-1}), (\Phi(\alpha_{i,t-1}))}{var(\Phi(\alpha_{i,t-1}))} > 0. \tag{10}$$

For the ease of interpretation, Table 12 reports $(\beta_{flow,performance}+1)/2$ which denotes the average likelihood that the sign of the fund flow [*S(flow)*] is positive (negative) conditional on the sign of the past performance measure [*S(performance)*] being positive (negative). The first inference from Table 12 is that all of the flow-performance sensitivity probability estimates, $(\beta_{flow,performance} + 1)/2$, are greater than 50% implying that a positive flow-performance relation exists for all of the different performance measures. Second, we find that the sensitivity is largest for the Sharpe ratio (the direction of the flow and past performance agrees 54.25% of the time for the Sharpe ratio, versus 54.18% for the raw return and 53.33% for the CAPM bond and stock alpha). It is noticeable that a significant fraction of flows remains unexplained when only considering past performance measures as none of the measures can explain more than 55%.

Furthermore, we can consider pairwise comparisons of two performance measures (models) and test which better captures how investors assess fund performance to allocate their capital by the following equation:

$$\Phi(Flow_{it}) = a + b_1\left(\frac{\Phi(\alpha^{m1}_{it-1})}{var(\Phi(\alpha^{m1}_{it-1}))} - \frac{\Phi(\alpha^{m2}_{it-1})}{var(\Phi(\alpha^{m2}_{it-1}))}\right) + \xi_{it}. \tag{11}$$

If the coefficient of this regression is positive (i.e., $b_1 > 0$), it implies that the flow-performance regression coefficient of model $m1$ is larger than that of model $m2$, and we can infer that model $m1$ better explains the subsequent fund flows than model $m2$.

Table 13 presents the results for the model comparison. The first two columns provide the coefficient estimate and double-clustered (by fund and month) t-statistic of the univariate regression of signed flows on signed out-performance (Equation 10). Columns 3 to 9 report the t-statistics of the pairwise test coefficients b_1 in Equation (11). Consistent with our main results, we find that the Sharpe ratio explains the flow performance sensitivity better than all factor models.[18]

V Implications

Style or factor investing strategies have been implemented widely in equity markets. However, there are currently few investment vehicles for investors to harvest factor premiums in the corporate bond market. Implementation of bond factors in investment portfolios may not be easy because bond trading costs can be very high. This suggests that bond investors may be less aware of factor models than investors of equity funds. Therefore, our finding that investors use simple measures instead of factor models to evaluate fund performance, may be not that surprising.

Corporate bonds are relatively more illiquid compared with government bonds and equities. The Investment Company Act of 1940 states that "A fund is generally required to price its portfolio using *readily available* market quotations". Most funds use each security's last traded price. In contrast to equities, the majority of bond trading takes place in over-the-counter dealer markets instead of on centralized exchanges. Thus, bond mutual funds do not have access to the same single exchange-determined closing prices as equity mutual funds. Moreover, many corporate bonds are held mainly as long-term investments in insurance companies or pension funds' portfolios and trade rarely after an initial distribution period. In

[18]Naturally, we cannot make out a significant difference between the Sharpe ratio and the raw return. This is because the Berk and Van Binsbergen (2016) approach only considers the sign of the performance measure. The sign of the Sharpe ratio is, to a large extent, defined by the sign of the return.

case of thinly-traded bonds, "the fund is permitted to value the securities at their fair value determined in good faith by, or under the direction of the fund's board of directors". However, there is much ambiguity about how the fund managers price those illiquid assets on a daily basis, which may lead to stale and inaccurate NAV calculations.

Cici et al. (2011) study the dispersion of month-end valuations related to bond-specific characteristics (associated with liquidity and market volatility) placed on identical corporate bonds by different mutual funds. Their tests reveal marking patterns that are consistent with returns smoothing behavior by managers. Funds with ambiguous marking policies and those holding "hard-to-mark" bonds appear more prone to smooth reported returns. Return smoothing distorts a fund's risk-return profile, perhaps leading investors to make suboptimal allocation decisions. Mutual fund managers compete with each other to attract new fund inflows on the basis of risk-adjusted performance statistics. Thus, our finding that investors mainly evaluate the fund managers by their funds' Sharpe ratios indicates that managers have strong incentives to smooth returns. Due to SEC regulations, managers of funds concentrating on US Treasury bond investments have little scope to shade their marks. However, corporate bond managers could have substantial room to adjust prices of their illiquid, thinly traded securities upward or downward to smooth returns.

Previous studies find evidence of large fund trading flows and large excess returns to stale price-oriented mutual fund trading strategies. Boudoukh et al. (2002), Goetzmann et al. (2001), Chalmers et al. (2001), Zitzewitz (2003) and Greene and Hodges (2002) show that mutual fund return predictability caused by stale prices allows profitable trading strategies in international and domestic equity funds as well as in high-yield US bond funds. Trade that exploits predictable fund returns results in a direct transfer of wealth from investors to active fund traders.

There is a mismatch between the illiquidity of corporate bond funds' underlying assets and the liquidity they offer to investors by providing the withdrawal rights on a daily basis. If most clients of corporate bond funds use the Sharpe ratio as the main fund performance measure, these investors are prone to non-optimal investment decisions caused by mismeasured fund performance. This pattern poses a threat to both buy-and-hold investors and managers of stale price oriented-funds because active traders can exploit the arbitrage opportunities by stale and mispriced NAV. On the one hand, return smoothing behavior of fund managers to boost and manipulate the Sharpe ratio misleads retail, less sophisticated, buy-and-hold investors to attract their flows into funds. On the other hand, this behavior creates trading opportunities for active traders. When the trading strategies are successfully implemented, buy-and-hold investors will suffer from the offsetting losses and expenses (for example from dilution effects) and fund managers are forced to sell good securities to have enough cash for redemptions.

VI Concluding Remarks

Which approach do investors use to measure performance in corporate bond mutual funds? To answer this question, we analyze the relation between mutual fund flows and different performance measures. We run a horse race among different performance measures ranging from the simple raw return and the Sharpe ratio to alphas estimated by using single and different multi-factor models. Our empirical analysis reveals that the net flows into actively managed U.S. corporate bond mutual funds are best explained by the Sharpe ratio. It thus seems that most investors do not use any factor model at all.

The use of the Sharpe ratio as performance measure is problematic for several reasons. First, a stale price effect is generally prevalent in bond funds, as documented in Qian (2011), Getmansky et al. (2004), Goldstein et al. (2017), Chen et al. (2010b) and Chen and Qin (2016). It facilitates opportunistic behavior of fund managers to boost the measured Sharpe ratio on purpose (for example by holding illiquid assets or 'hard-to-mark' bonds), especially given that investors appear to evaluate fund performance by using the Sharpe ratio. Therefore, our findings have potentially important implications for both investors and manager of corporate bond mutual funds. We suggest that investors should not use the Sharpe ratio without the adjustment for the effect of smoothing returns and investors should be cautious to funds' manipulation of reported measures.

Moreover, mutual fund return predictability caused by stale prices allows profitable trading strategies (Boudoukh et al., 2002; Goetzmann et al., 2001; Chalmers et al., 2001; Zitzewitz, 2003; Greene and Hodges, 2002 document evidence in international and domestic equity funds as well as in high-yield US bond funds). Gains earned by active fund traders from trades that exploit fund net asset value (NAV) that reflect stale prices are matched with the expenses and losses suffered by buy-and-hold fund investors who are normally less sophisticated (i.e., retail investors, individual clients purchasing fund share though broker-sold distribution channel). Fund managers of those funds, especially no-load funds, may be forced to sell good assets to raise cash for redemptions.

Further studies are needed to find out a "fair pricing" mechanism to address the dispersion in corporate bond valuation across mutual funds to align the NAV with the "true" value of a fund share. This would help avoid wealth transfers among existing, new and redeeming fund investors. Furthermore, we believe that further research should be undertaken to explore bond factors and make bond factor investing strategies more feasible in the future, so that corporate bond mutual funds can successfully adopt and provide vehicles for bond investors to harvest factor premiums in corporate bond markets.

References

Agarwal, V., Green, T. C., and Ren, H. (2018). Alpha or beta in the eye of the beholder: What drives hedge fund flows? *Journal of Financial Economics*, 127(3):417–434.

Aragon, G. O., Li, L., and Qian, J. (2019). The use of credit default swaps by bond mutual funds: Liquidity provision and counterparty risk. *Journal of Financial Economics*, 131(1):168–185.

Avramov, D., Chordia, T., Jostova, G., and Philipov, A. (2007). Momentum and credit rating. *Journal of Finance*, 62(5):2503–2520.

Bai, J., Bali, T. G., and Wen, Q. (2019). Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, 131(3):619–642.

Barber, B. M., Huang, X., and Odean, T. (2016). Which factors matter to investors? Evidence from mutual fund flows. *Review of Financial Studies*, 29(10):2600–2642.

Bektic, D., Wenzler, J.-S., Wegener, M., Schiereck, D., and Spielmann, T. (2017). Extending Fama-French factors to corporate bond markets. *Forthcoming in the Journal of Portfolio Management.*

Ben-David, I., Li, J., Rossi, A., and Song, Y. (2019). What do mutual fund investors really care about? *Fisher College of Business Working Paper*, (2019-03):005.

Berk, J. B. and Van Binsbergen, J. H. (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics*, 118(1):1–20.

Berk, J. B. and Van Binsbergen, J. H. (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics*, 119(1):1–23.

Blocher, J. and Molyboga, M. (2017). The revealed preference of sophisticated investors. *European Financial Management*, 23(5):839–872.

Bollen, N. P. and Pool, V. K. (2008). Conditional return smoothing in the hedge fund industry. *Journal of Financial and Quantitative Analysis*, 43(2):267–298.

Bollen, N. P. and Pool, V. K. (2009). Do hedge fund managers misreport returns? Evidence from the pooled distribution. *Journal of Finance*, 64(5):2257–2288.

Boudoukh, J., Richardson, M., Subrahmanyam, M., and Whitelaw, R. F. (2002). Stale prices and strategies for trading mutual funds. *Financial Analysts Journal*, 58(4):53–71.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82.

Chakraborty, I., Kumar, A., Muhlhofer, T., and Sastry, R. (2018). If you show it, they will use it: The'capm heuristic'and mutual fund allocation decisions. *Working paper, University of Miami.*

Chalmers, J., Edelen, R., and Kadlec, G. (2001). On the perils of security pricing by financial intermediaries: The wildcard option in transacting mutual fund shares. *Journal of Finance*, 56(6):2209–36.

Chen, Q., Goldstein, I., and Jiang, W. (2010a). Payoff complementarities and financial fragility: Evidence from mutual fund outflows. *Journal of Financial Economics*, 97(2):239–262.

Chen, Y., Ferson, W., and Peters, H. (2010b). Measuring the timing ability and performance of bond mutual funds. *Journal of Financial Economics*, 98(1):72–89.

Chen, Y. and Qin, N. (2016). The behavior of investor flows in corporate bond mutual funds. *Management Science*, 63(5):1365–1381.

Chung, K. H., Wang, J., and Wu, C. (2019). Volatility and the cross-section of corporate bond returns. *Journal of Financial Economics*.

Cici, G., Gibson, S., and Merrick Jr, J. J. (2011). Missing the marks? Dispersion in corporate bond valuations across mutual funds. *Journal of Financial Economics*, 101(1):206–226.

Cochrane, J. H. and Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, 95(1):138–160.

Comer, G. and Rodriguez, J. (2013). A comparison of corporate versus government bond funds. *Journal of Economics and Finance*, 37(4):495–510.

Del Guercio, D. and Tkac, P. A. (2008). Star power: The effect of monrningstar ratings on mutual fund flow. *Journal of Financial and Quantitative Analysis*, 43(4):907–936.

Elton, E. J. and Gruber, M. J. (2013). Mutual funds. In *Handbook of the Economics of Finance*, volume 2, pages 1011–1061. Elsevier.

Elton, E. J., Gruber, M. J., Agrawal, D., and Mann, C. (2001). Explaining the rate spread on corporate bonds. *Journal of Finance*, 56(1):247–277.

Elton, E. J., Gruber, M. J., and Blake, C. R. (1995). Fundamental economic variables, expected returns, and bond fund performance. *Journal of Finance*, 50(4):1229–1256.

Evans, R. B. (2010). Mutual fund incubation. *Journal of Finance*, 65(4):1581–1611.

Evans, R. B. and Sun, Y. (2018). Models or stars: The role of asset pricing models and heuristics in investor risk adjustment. *Working paper, University of Virginia.*

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

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Fulkerson, J. A., Jordan, B. D., and Riley, T. B. (2013). Return chasing in bond funds. *Journal of Fixed Income*, 22(4):90.

Fung, W. and Hsieh, D. A. (2004). Hedge fund benchmarks: A risk-based approach. *Financial Analysts Journal*, 60(5):65–80.

Gebhardt, W. R., Hvidkjaer, S., and Swaminathan, B. (2005a). The cross-section of expected corporate bond returns: Betas or characteristics? *Journal of Financial Economics*, 75(1):85–114.

Gebhardt, W. R., Hvidkjaer, S., and Swaminathan, B. (2005b). Stock and bond market interaction: Does momentum spill over? *Journal of Financial Economics*, 75(3):651–690.

Getmansky, M., Lo, A. W., and Makarov, I. (2004). An econometric model of serial correlation and illiquidity in hedge fund returns. *Journal of Financial Economics*, 74(3):529–609.

Goetzmann, W., Ingersoll, J., Spiegel, M., and Welch, I. (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies*, 20(5):1503–1546.

Goetzmann, W. N., Ivković, Z., and Rouwenhorst, K. G. (2001). Day trading international mutual funds: Evidence and policy solutions. *Journal of Financial and Quantitative Analysis*, 36(3):287–309.

Goldstein, I., Jiang, H., and Ng, D. T. (2017). Investor flows and fragility in corporate bond funds. *Journal of Financial Economics*, 126(3):592–613.

Greene, J. T. and Hodges, C. W. (2002). The dilution impact of daily fund flows on open-end mutual funds. *Journal of Financial Economics*, 65(1):131–158.

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, 28(3):650–705.

Houweling, P. and Van Zundert, J. (2017). Factor investing in the corporate bond market. *Financial Analysts Journal*, 73(2):100–115.

Huang, J., Sialm, C., and Zhang, H. (2011). Risk shifting and mutual fund performance. *Review of Financial Studies*, 24(8):2575–2616.

Huij, J. and Derwall, J. (2008). "hot hands" in bond funds. *Journal of Banking & Finance*, 32(4):559–572.

Israel, R., Palhares, D., and Richardson, S. (2018). Common factors in corporate bond returns. *Journal of Investment Management*, 16(2).

Jiang, G. J. and Yuksel, H. Z. (2017). What drives the "smart-money" effect? Evidence from investors' money flow to mutual fund classes. *Journal of Empirical Finance*, 40:39–58.

Jordan, B. D. and Riley, T. B. (2015). Volatility and mutual fund manager skill. *Journal of Financial Economics*, 118(2):289–298.

Jostova, G., Nikolova, S., Philipov, A., and Stahel, C. W. (2013). Momentum in corporate bond returns. *Review of Financial Studies*, 26(7):1649–1693.

Koijen, R. S., Lustig, H., and Van Nieuwerburgh, S. (2017). The cross-section and time series of stock and bond returns. *Journal of Monetary Economics*, 88:50–69.

Lin, H., Wang, J., and Wu, C. (2011). Liquidity risk and expected corporate bond returns. *Journal of Financial Economics*, 99(3):628–650.

Ludvigson, S. C. and Ng, S. (2009). Macro factors in bond risk premia. *Review of Financial Studies*, 22(12):5027–5067.

Pástor, L. and Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3):642–685.

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2015). Scale and skill in active management. *Journal of Financial Economics*, 116(1):23–45.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, 22(1):435–480.

Qian, M. (2011). Stale prices and the performance evaluation of mutual funds. *Journal of Financial and Quantitative Analysis*, 46(2):369–394.

Stambaugh, R. F. and Yuan, Y. (2016). Mispricing factors. *Review of Financial Studies*, 30(4):1270–1315.

Wermers, R. (2011). Performance measurement of mutual funds, hedge funds, and institutional accounts. *Annual Review of Financial Economics*, 3(1):537–574.

Zhao, X. (2005). Determinants of flows into retail bond funds. *Financial Analysts Journal*, 61(4):47–59.

Zitzewitz, E. (2003). Who cares about shareholders? Arbitrage-proofing mutual funds. *Journal of Law, Economics, and Organization*, 19(2):245–280.

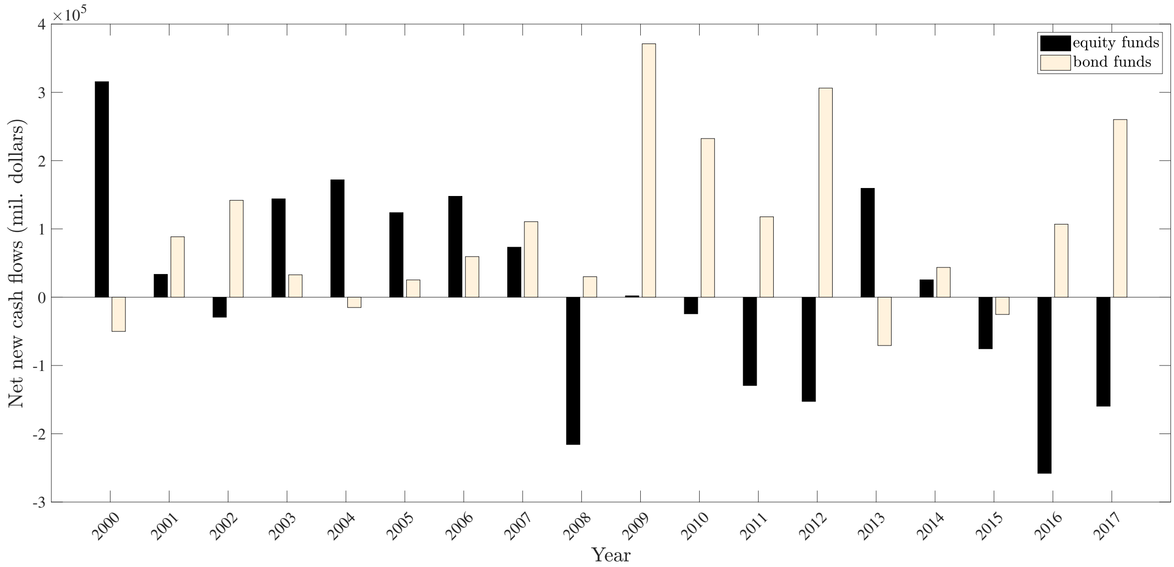

Figure 1: Net Cash Flows Through Time

The figure plots the net cash flows into equity and bond mutual funds over the period 2000 to 2017. The data are from the Investment Company Institute Fact Book 2018.

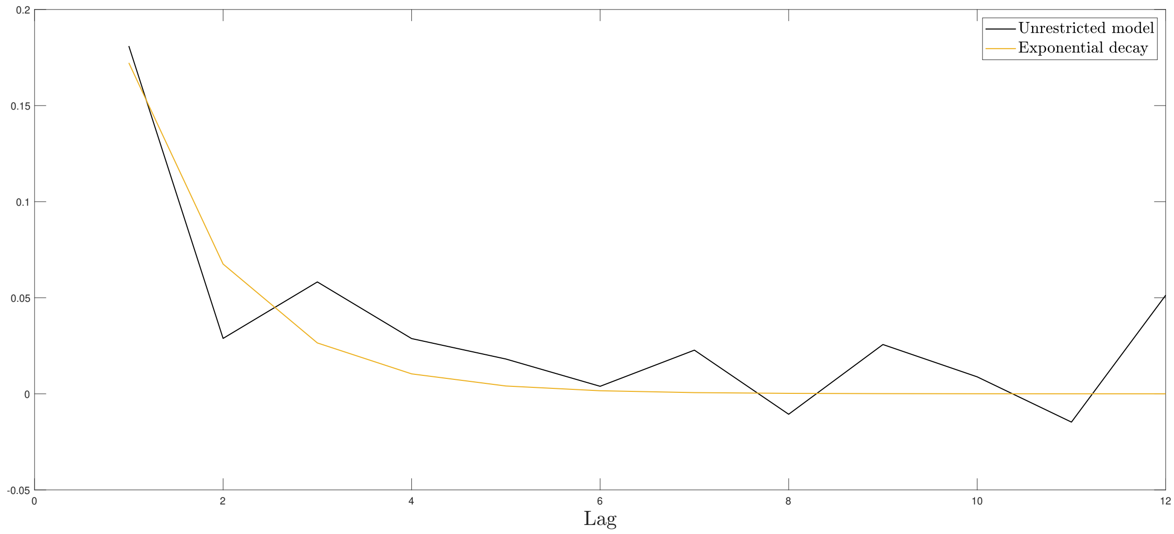

Figure 2: Fund Flow–Past Return Relation

The graph presents the regression coefficient estimates (y axis) at various lags (x axis) for two models of monthly fund flows: (1) a simple unrestricted model (with twelve lags of monthly fund returns and individual coefficient estimates on each lagged return) and (2) an exponential decay model as in Equation (6) with the decay rate parameter λ.

Table 1: Descriptive Statistics

This table presents summary statistics. Our sample contains 3,647 unique fund share classes (1060 high yield bond share classes and 2587 investment grade share classes) of actively managed U.S. corporate bond mutual funds from January 1996 to June 2017. Panel A summarizes the fund characteristics. SD denotes the standard deviation. P25 and P75 are the 25% and 75% quantiles, respectively. $\rho 1$ is the first-order autocorrelation, reported in percentage points. Panel B reports the time-series average of the cross-sectional distribution of fund returns and flows. Panel C summarizes the time-series of the different model alphas estimated using rolling 60-month fund past returns. Panel D presents the correlation matrix between different performance measures. The unit of observation is fund share-month.

Panel A: Fund characteristics

	Mean	SD	P25	Median	P75
Size ($mil)	609.15	2913.23	39.10	113.00	376.70
Age (years)	10.61	8.90	4.42	8.59	14.39
Expense ratio (%)	0.96	0.47	0.62	0.85	1.22
Noload dummy	0.32	0.47	0.00	0.00	1.00
Volatility (t-12 to t-1)(%)	1.22	1.03	0.68	0.96	1.44

Panel B: Fund return and flow

	Mean	SD	P25	Median	P75	$\rho 1$
	Fund return (% per month)					
All funds	0.42	1.65	-0.15	0.43	1.12	18.14
Investment grade	0.37	1.22	-0.12	0.36	0.94	16.15
High yield	0.55	2.41	-0.31	0.72	1.62	23.07
	Fund flow (% per month)					
All funds	0.58	5.09	-1.57	-0.11	1.76	28.32
Investment grade	0.59	4.94	-1.47	-0.07	1.72	28.39
High yield	0.57	5.44	-1.81	-0.23	1.87	28.17

Panel C: Fund alpha

	Mean	SD	P25	Median	P75
CAPMbond alpha	0.062	1.538	-0.207	0.032	0.388
CAPMbondstk alpha	-0.044	1.133	-0.286	-0.031	0.221
E4 alpha	-0.043	0.741	-0.229	-0.026	0.181
F5 alpha	-0.010	0.837	-0.303	-0.006	0.302
F7 alpha	-0.007	0.860	-0.306	-0.005	0.312

Panel D: Correlation between different performance measures

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
(a) Raw return	1.00	0.74	0.88	0.62	0.29	0.34	0.33	0.06
(b) Sharpe ratio		1.00	0.53	0.37	0.22	0.29	0.28	0.08
(c) CAPMbond alpha			1.00	0.69	0.34	0.36	0.35	0.05
(d) CAPMbondstk alpha				1.00	0.52	0.51	0.50	0.06
(e) E4 alpha					1.00	0.85	0.82	0.09
(f) F5 alpha						1.00	0.97	0.08
(g) F7 alpha							1.00	0.08
(h) MS rating								1.00

Table 2: Model Horse Race – Full Sample

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using the full corporate bond fund sample. We estimate the relation between flow and a fund's decile ranking based on different performance measures by running the following regression:

$$F_{pt} = a + \sum_{i} \sum_{j} b_{ij} D_{ij,p,t-1} + cX_{pt-1} + \mu_t + e_{pt},$$

F_{pt} is the fund flow of mutual fund p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund p in month $t-1$ is in the decile i based on the first model and decile j based on the second model. To estimate the model, the dummy variable for $i = 5$ and $j = 5$ is excluded. The matrix $X_{p,t-1}$ represents control variables including the lagged fund flow from month $t-1$, the lagged fund expense ratio, a dummy for no-load funds, a fund's return standard deviation estimated over the prior twelve months, the log of fund size and the log of fund age in month $t-1$ as well as Morningstar rating dummies in month $t-1$. We also include time fixed effects (μ_t).
For each pairwise comparison of two performance measures, we have 45 such b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji} to determine whether investors are more sensitive to the first or to the second measure (alpha estimated using the first model or using the second model), we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. If investors are more responsive to the first measure than the second one, we would expect to reject the null hypothesis in favor of the alternative hypothesis that b_{ij} is greater than b_{ji}. The table reports the results of two hypothesis tests for each horse race: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of difference equals 50%. We test the first hypothesis with a Wald test and the second one with a Binomial test. We present a model as "winning model" if the sum of coefficient differences is statistically significant from zero. A "–" indicates no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 2: Model Horse Race – Full Sample (continued)

A. Sharpe ratio

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	17.42***	14.16***	9.848***	6.547***	9.482***	9.462***
Wald p-Value	0.000	0.000	0.000	0.001	0.000	0.000
% of coeff. Diff. >0	91.11***	93.33***	88.89***	73.33***	84.44***	86.67***
Binomial p-Value	0.000	0.000	0.000	0.001	0.000	0.000

B. Raw return

Winning model	–	–	**E4**	–	–
Losing model	–	–	**RR**	–	–
	RR_Cb	RR_Csb	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-5.286	-4.041	-5.508**	-1.456	-0.994
Wald p-Value	0.154	0.1072	0.016	0.4901	0.6278
% of coeff. Diff. >0	28.89***	31.11***	15.56***	35.56**	40
Binomial p-Value	0.003	0.008	0.000	0.036	0.1163

C. CAPM bond

Winning model	–	**E4**	–	–
Losing model	–	**Cb**	–	–
	Cb_Csb	Cb_E4	Cb_F5	Cb_F7
Sum of coeff. Diff.	-2.308	-4.626**	0.180	0.936
Wald p-Value	0.373	0.046	0.938	0.675
% of coeff. Diff. >0	40.000	20.00***	57.780	53.330
Binomial p-Value	0.116	0.000	0.186	0.383

D. CAPM stock + bond

Winning model	**E4**	–	–
Losing model	**Csb**	–	–
	Csb_E4	Csb_F5	Csb_F7
Sum of coeff. Diff.	-4.841*	3.100	3.692
Wald p-Value	0.089	0.235	0.139
% of coeff. Diff. >0	42.220	62.22*	71.11***
Binomial p-Value	0.186	0.068	0.003

E. E4

Winning model	**E4**	**E4**
Losing model	**F5**	**F7**
	E4_F5	E4_F7
Sum of coeff. Diff.	7.710***	7.593***
Wald p-Value	0.006	0.002
% of coeff. Diff. >0	75.56***	86.67***
Binomial p-Value	0.000	0.000

F. F5

Winning model	–
Losing model	–
	F5_F7
Sum of coeff. Diff.	3.569
Wald p-Value	0.456
% of coeff. Diff. >0	53.330
Binomial p-Value	0.383

Table 3: Model Horse Race – High Yield Bond Funds

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using high yield bond fund sample. We estimate the relation between flow and a fund's decile ranking based on different performance measures by running the following regression:

$$F_{pt} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{pt-1} + \mu_t + e_{pt},$$

F_{pt} is the fund flow of mutual fund p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund p in month $t-1$ is in the decile i based on the first model and decile j based on the second model. To estimate the model, the dummy variable for $i = 5$ and $j = 5$ is excluded. The matrix $X_{p,t-1}$ represents control variables including the lagged fund flow from month $t-1$, the lagged fund expense ratio, a dummy for no-load funds, a fund's return standard deviation estimated over the prior twelve months, the log of fund size and the log of fund age in month $t-1$ as well as Morningstar rating dummies in month $t-1$. We also include time fixed effects (μ_t).

For each pairwise comparison of two performance measures, we have 45 such b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji} to determine whether investors are more sensitive to the first or to the second measure (alpha estimated using the first model or using the second model), we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. If investors are more responsive to the first measure than the second one, we would expect to reject the null hypothesis in favor of the alternative hypothesis that b_{ij} is greater than b_{ji}. The table reports the results of two hypothesis tests for each horse race: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of difference equals 50%. We test the first hypothesis with a Wald test and the second one with a Binomial test. We present a model as "winning model" if the sum of coefficient differences is statistically significant from zero. A "–" indicates no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 3: Model Horse Race – High Yield Bond Funds (continued)

A. Sharpe Ratio

Winning model **Losing model**	**SR** **RR**	**SR** **Cb**	**SR** **Csb**	**SR** **E4**	**SR** **F5**	**SR** **F7**
	SR_RR	SR_Cb	SR_Csb	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	22.53***	21.85***	20.80***	10.11***	11.69***	9.514***
Wald p-Value	0.00	0.00	0.00	0.01	0.00	0.001
% of coeff. Diff. >0	86.67***	93.33***	93.33***	68.89***	82.22***	77.78***
Binomial p-Value	0.00	0.00	0.00	0.01	0.00	0.00

B. Raw return

Winning model **Losing model**	– –	– –	**E4** **RR**	**F5** **RR**	**F7** **RR**
	RR_Cb	RR_Csb	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	8.599	5.455	-6.335*	-5.862*	-5.743*
Wald p-Value	0.2948	0.1368	0.094	0.094	0.076
% of coeff. Diff. >0	66.67**	55.56	31.11***	33.33**	44.44
Binomial p-Value	0.02	0.2757	0.01	0.02	0.2757

C. CAPM bond

Winning model **Losing model**	– –	**E4** **Cb**	**F5** **Cb**	**F7** **Cb**
	Cb_Csb	Cb_E4	Cb_F5	Cb_F7
Sum of coeff. Diff.	2.946	-10.70**	-8.893**	-8.563**
Wald p-Value	0.451	0.01	0.036	0.028
% of coeff. Diff. >0	51.110	22.22***	20.00***	33.33**
Binomial p-Value	0.500	0.00	0.00	0.02

D. CAPM stock + bond

Winning model **Losing model**	**E4** **Csb**	**F5** **Csb**	**F7** **Csb**
	Csb_E4	Csb_F5	Csb_F7
Sum of coeff. Diff.	-20.02***	-12.31**	-11.93**
Wald p-Value	0.00	0.04	0.02
% of coeff. Diff. >0	8.889***	24.44***	24.44***
Binomial p-Value	0.000	0.00	0.00

E. E4

Winning model **Losing model**	– –	– –
	E4_F5	E4_F7
Sum of coeff. Diff.	5.937	3.827
Wald p-Value	0.315	0.421
% of coeff. Diff. >0	62.22*	55.560
Binomial p-Value	0.07	0.276

F. F5

Winning model **Losing model**	– –
	F5_F7
Sum of coeff. Diff.	-3.841
Wald p-Value	0.619
% of coeff. Diff. >0	42.220
Binomial p-Value	0.186

Table 4: Model Horse Race – Investment Grade Bond Funds

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using investment grade bond fund sample. We estimate the relation between flow and a fund's decile ranking based on different performance measures by running the following regression:

$$F_{pt} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{pt-1} + \mu_t + e_{pt},$$

F_{pt} is the fund flow of mutual fund p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund p in month $t-1$ is in the decile i based on the first model and decile j based on the second model. To estimate the model, the dummy variable for $i = 5$ and $j = 5$ is excluded. The matrix $X_{p,t-1}$ represents control variables including the lagged fund flow from month $t-1$, the lagged fund expense ratio, a dummy for no-load funds, a fund's return standard deviation estimated over the prior twelve months, the log of fund size and the log of fund age in month $t-1$ as well as Morningstar rating dummies in month $t-1$. We also include time fixed effects (μ_t).
For each pairwise comparison of two performance measures, we have 45 such b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji} to determine whether investors are more sensitive to the first or to the second measure (alpha estimated using the first model or using the second model), we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. If investors are more responsive to the first measure than the second one, we would expect to reject the null hypothesis in favor of the alternative hypothesis that b_{ij} is greater than b_{ji}. The table reports the results of two hypothesis tests for each horse race: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of difference equals 50%. We test the first hypothesis with a Wald test and the second one with a Binomial test. We present a model as "winning model" if the sum of coefficient differences is statistically significant from zero. A "–" indicates no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 4: Model Horse Race – Investment Grade Bond Funds (continued)

A. Sharpe Ratio

Winning model **Losing model**	**SR** **RR**	**SR** **Cb**	**SR** **Csb**	**SR** **E4**	**SR** **F5**	**SR** **F7**
	SR_RR	SR_Cb	SR_Csb	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	13.43***	7.173**	4.325*	4.735**	7.870***	8.310***
Wald p-Value	0.00	0.015	0.056	0.014	0.000	0.000
% of coeff. Diff. >0	84.44***	82.22***	64.44**	68.89***	75.56***	82.22***
Binomial p-Value	0.00	0.00	0.04	0.01	0.00	0.00

B. Raw return

Winning model **Losing model**	**Cb** **RR**	**Csb** **RR**	**E4** **RR**	– –	– –
	RR_Cb	RR_Csb	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-7.388**	-5.865**	-4.702**	-0.12	0.271
Wald p-Value	0.030	0.024	0.043	0.9553	0.8983
% of coeff. Diff. >0	26.67***	26.67***	33.33**	46.67	40
Binomial p-Value	0.00	0.00	0.02	0.383	0.1163

C. CAPM bond

Winning model **Losing model**	– –	– –	**Cb** **F5**	**Cb** **F7**
	Cb_Csb	Cb_E4	Cb_F5	Cb_F7
Sum of coeff. Diff.	-3.069	-0.684	4.027*	4.682**
Wald p-Value	0.267	0.770	0.080	0.037
% of coeff. Diff. >0	40.000	46.670	71.11***	71.11***
Binomial p-Value	0.116	0.383	0.00	0.00

D. CAPM stock + bond

Winning model **Losing model**	– –	**Csb** **F5**	**Csb** **F7**
	Csb_E4	Csb_F5	Csb_F7
Sum of coeff. Diff.	1.240	6.773***	7.576***
Wald p-Value	0.654	0.008	0.003
% of coeff. Diff. >0	53.330	82.22***	82.22***
Binomial p-Value	0.383	0.00	0.00

E. E4

Winning model **Losing model**	**E4** **F5**	**E4** **F7**
	E4_F5	E4_F7
Sum of coeff. Diff.	6.685**	8.073***
Wald p-Value	0.010	0.001
% of coeff. Diff. >0	71.11***	82.22***
Binomial p-Value	0.00	0.00

F. F5

Winning model **Losing model**	– –
	F5_F7
Sum of coeff. Diff.	8.119
Wald p-Value	0.138
% of coeff. Diff. >0	55.560
Binomial p-Value	0.276

Table 5: Model Horse Race – Retail-Oriented Bond Funds

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using retail-oriented bond fund sample. We estimate the relation between flow and a fund's decile ranking based on different performance measures by running the following regression:

$$F_{pt} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{pt-1} + \mu_t + e_{pt},$$

F_{pt} is the fund flow of mutual fund p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund p in month $t-1$ is in the decile i based on the first model and decile j based on the second model. To estimate the model, the dummy variable for $i = 5$ and $j = 5$ is excluded. The matrix $X_{p,t-1}$ represents control variables including the lagged fund flow from month $t-1$, the lagged fund expense ratio, a dummy for no-load funds, a fund's return standard deviation estimated over the prior twelve months, the log of fund size and the log of fund age in month $t-1$ as well as Morningstar rating dummies in month $t-1$. We also include time fixed effects (μ_t).

For each pairwise comparison of two performance measures, we have 45 such b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji} to determine whether investors are more sensitive to the first or to the second measure (alpha estimated using the first model or using the second model), we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. If investors are more responsive to the first measure than the second one, we would expect to reject the null hypothesis in favor of the alternative hypothesis that b_{ij} is greater than b_{ji}. The table reports the results of two hypothesis tests for each horse race: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of difference equals 50%. We test the first hypothesis with a Wald test and the second one with a Binomial test. We present a model as "winning model" if the sum of coefficient differences is statistically significant from zero. A "–" indicates no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 5: Model Horse Race – Retail-Oriented Bond Funds (continued)

A. Sharpe ratio

Winning model **Losing model**	**SR** **RR**	**SR** **Cb**	**SR** **Csb**	**SR** **E4**	**SR** **F5**	**SR** **F7**
	SR12_RR	SR12_Cb	SR12_Csb	SR12_E4	SR12_F5	SR12_F7
Sum of coeff. Diff.	15.19***	13.16***	12.78***	9.043***	10.48***	10.41***
Wald p-Value	0.000	0.000	0.000	0.000	0.000	0.000
% of coeff. Diff. >0	91.11***	93.33***	95.56***	80.00***	86.67***	82.22***
Binomial p-Value	0.000	0.000	0.000	0.000	0.000	0.000

B. Raw return

Winning model **Losing model**	– –	– –	– –	– –	– –
	RR_Cb	RR_Csb	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-2.805	0.725	-2.07	1.086	1.508
Wald p-Value	0.462	0.78	0.396	0.619	0.47
% of coeff. Diff. >0	44.44	55.56	37.78*	60	53.33
Binomial p-Value	0.276	0.276	0.068	0.116	0.383

C. CAPM bond

Winning model **Losing model**	– –	– –	– –	– –
	Cb_Csb	Cb_E4	Cb_F5	Cb_F7
Sum of coeff. Diff.	3.274	-1.251	2.312	2.717
Wald p-Value	0.230	0.598	0.296	0.195
% of coeff. Diff. >0	60.000	46.670	60.000	64.44**
Binomial p-Value	0.116	0.383	0.116	0.036

D. CAPM stock + bond

Winning model **Losing model**	– –	– –	– –
	Csb_E4	Csb_F5	Csb_F7
Sum of coeff. Diff.	-4.347	1.575	2.140
Wald p-Value	0.120	0.512	0.350
% of coeff. Diff. >0	42.220	60.000	57.780
Binomial p-Value	0.186	0.116	0.186

E. E4

Winning model **Losing model**	**E4** **F5**	**E4** **F7**
	E4_F5	E4_F7
Sum of coeff. Diff.	6.640**	6.133**
Wald p-Value	0.020	0.013
% of coeff. Diff. >0	71.11***	71.11***
Binomial p-Value	0.003	0.003

F. F5

Winning model **Losing model**	– –
	F5_F7
Sum of coeff. Diff.	2.805
Wald p-Value	0.601
% of coeff. Diff. >0	51.110
Binomial p-Value	0.500

Table 6: Model Horse Race – Institutional-Oriented Bond Funds

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using institutional-oriented bond fund sample. We estimate the relation between flow and a fund's decile ranking based on different performance measures by running the following regression:

$$F_{pt} = a + \sum_{i}\sum_{j} b_{ij} D_{ij,p,t-1} + cX_{pt-1} + \mu_t + e_{pt},$$

F_{pt} is the fund flow of mutual fund p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund p in month $t-1$ is in the decile i based on the first model and decile j based on the second model. To estimate the model, the dummy variable for $i = 5$ and $j = 5$ is excluded. The matrix $X_{p,t-1}$ represents control variables including the lagged fund flow from month $t-1$, the lagged fund expense ratio, a dummy for no-load funds, a fund's return standard deviation estimated over the prior twelve months, the log of fund size and the log of fund age in month $t-1$ as well as Morningstar rating dummies in month $t-1$. We also include time fixed effects (μ_t).
For each pairwise comparison of two performance measures, we have 45 such b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji} to determine whether investors are more sensitive to the first or to the second measure (alpha estimated using the first model or using the second model), we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. If investors are more responsive to the first measure than the second one, we would expect to reject the null hypothesis in favor of the alternative hypothesis that b_{ij} is greater than b_{ji}. The table reports the results of two hypothesis tests for each horse race: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of difference equals 50%. We test the first hypothesis with a Wald test and the second one with a Binomial test. We present a model as "winning model" if the sum of coefficient differences is statistically significant from zero. A "–" indicates no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 6: Model Horse Race – Institutional-Oriented Bond Funds (continued)

A. Sharpe Ratio

Winning model	**SR**	**SR**	–	–	**SR**	**SR**
Losing model	**RR**	**Cb**	–	–	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	24.32***	18.50***	4.967	3.745	8.867***	7.642**
Wald p-Value	0.00	0.01	0.2328	0.2752	0.007	0.017
% of coeff. Diff. >0	86.67***	75.56***	64.44**	57.78	75.56***	64.44**
Binomial p-Value	0.00	0.00	0.04	0.1856	0.00	0.04

B. Raw return

Winning model	**Cb**	**Csb**	**E4**	**F5**	**F7**
Losing model	**RR**	**RR**	**RR**	**RR**	**RR**
	RR_Cb	RR_Csb	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-11.09*	-13.61***	-12.25***	-7.679**	-7.672**
Wald p-Value	0.06	0.00	0.00	0.036	0.038
% of coeff. Diff. >0	40	28.89***	28.89***	37.78*	31.11***
Binomial p-Value	0.1163	0.00	0.00	0.07	0.01

C. CAPM bond

Winning model	**Csb**	**E4**	–	–
Losing model	**Cb**	**Cb**	–	–
	Cb_Csb	Cb_E4	Cb_F5	Cb_F7
Sum of coeff. Diff.	-16.57***	-10.04*	-4.904	-4.789
Wald p-Value	0.00	0.06	0.322	0.317
% of coeff. Diff. >0	17.78***	26.67***	40.000	44.440
Binomial p-Value	0.00	0.00	0.116	0.276

D. CAPM stock + bond

Winning model	–	–	–
Losing model	–	–	–
	Csb_E4	Csb_F5	Csb_F7
Sum of coeff. Diff.	-6.354	4.787	3.584
Wald p-Value	0.323	0.377	0.501
% of coeff. Diff. >0	48.890	66.67**	53.330
Binomial p-Value	0.500	0.02	0.383

E. E4

Winning model	–	–
Losing model	–	–
	E4_F5	E4_F7
Sum of coeff. Diff.	4.054	5.006
Wald p-Value	0.462	0.299
% of coeff. Diff. >0	62.22*	71.11***
Binomial p-Value	0.07	0.00

F. F5

Winning model	–
Losing model	–
	F5_F7
Sum of coeff. Diff.	-2.707
Wald p-Value	0.816
% of coeff. Diff. >0	46.670
Binomial p-Value	0.383

Table 7: Response of Fund Flows to Different Components of Fund Returns

This table reports the coefficient estimates from panel regressions of percentage fund flows on the components of a fund's return based on two models: the Elton et al. (1995) four-factor model (Panel A) and a seven-factor model which is an augmented Fama and French (1993) five-factor model added momentum and liquidity factors (Panel B). The regressions also include control variables and month fixed effects. The standard errors are double-clustered by fund and month (p-Values are in parentheses). *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

A. Elton et al. (1995) Four-Factor Model

	All funds	Investment grade	High yield
ALPHA	0.121***	0.207***	0.102*
	(0.002)	(0.000)	(0.050)
MKTRET stk	0.032	0.034	0.017
	(0.435)	(0.455)	(0.785)
MKTRET bond	0.080	0.019	0.394**
	(0.253)	(0.761)	(0.023)
DEFRET	−0.061	0.058	−0.155***
	(0.186)	(0.257)	(0.005)
OPTIONRET	0.167*	0.151	−0.010
	(0.092)	(0.181)	(0.938)

B. Seven-Factor Model

	All funds	Investment grade	High yield
ALPHA	0.144***	0.238***	0.118**
	(0.001)	(0.000)	(0.023)
MKTRET stk	0.034	0.051	−0.021
	(0.382)	(0.242)	(0.717)
SIZRET	0.160	0.253	0.137
	(0.156)	(0.113)	(0.359)
VALRET	0.194*	0.310**	0.138
	(0.061)	(0.017)	(0.221)
MOMRET	0.105	0.286***	0.093
	(0.198)	(0.004)	(0.407)
TERMRET	0.127**	0.112**	0.453***
	(0.026)	(0.036)	(0.000)
DEFRET	−0.070	0.045	−0.174***
	(0.136)	(0.360)	(0.002)
LIQRET	0.187	0.146	0.173
	(0.198)	(0.491)	(0.185)

Table 8: Model Horse Race (12-Month Window)

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using full bond fund sample. We estimate the relation between flow and a fund's decile ranking based on different performance measures by running the following regression:

$$F_{pt} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{pt-1} + \mu_t + e_{pt},$$

$F_{pt}$is the fund flow of mutual fund p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund p in month $t-1$ is in the decile i based on the first measure and decile j based on the second measure. Each of measure is calculated as a weighted average of the prior twelve monthly alphas (or returns for the Sharpe ratio). To estimate the model, the dummy variable for $i = 5$ and $j = 5$ is excluded. The matrix $X_{p,t-1}$ represents control variables including the lagged fund flow from month $t-1$, the lagged fund expense ratio, a dummy for no-load funds, a fund's return standard deviation estimated over the prior twelve months, the log of fund size and the log of fund age in month $t-1$ as well as Morningstar rating dummies in month $t-1$. We also include time fixed effects (μ_t).

For each pairwise comparison of two performance measures, we have 45 such b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji} to determine whether investors are more sensitive to the first or to the second measure (alpha estimated using the first model or using the second model), we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. If investors are more responsive to the first measure than the second one, we would expect to reject the null hypothesis in favor of the alternative hypothesis that b_{ij} is greater than b_{ji}. The table reports the results of two hypothesis tests for each horse race: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of difference equals 50%. We test the first hypothesis with a Wald test and the second one with a Binomial test. We present a model as "winning model" if the sum of coefficient differences is statistically significant from zero. A "–" indicates no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 8: Model Horse Race (12-Month Window) (continued)

A. Weighted Average Sharpe Ratio

Winning model **Losing model**	**SR** **RR**	**SR** **Cb**	**SR** **Csb**	**SR** **E4**	**SR** **F5**	**SR** **F7**
	SR_RR	SR_Cb	SR_Csb	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	14.43***	5.664***	4.369**	2.807*	6.098***	6.911***
Wald p-Value	0.00	0.009	0.016	0.071	0.000	0.000
% of coeff. Diff. >0	84.44***	68.89***	60	53.33	84.44***	82.22***
Binomial p-Value	0.00	0.01	0.1163	0.383	0.00	0.00

B. Raw return

Winning model **Losing model**	**Cb** **RR**	**Csb** **RR**	**E4** **RR**		
	RR_Cb	RR_Csb	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-9.208***	-5.482***	-5.842***	-1.469	-0.758
Wald p-Value	0.003	0.008	0.002	0.4107	0.6627
% of coeff. Diff. >0	26.67***	22.22***	26.67***	37.78*	46.67
Binomial p-Value	0.00	0.00	0.00	0.07	0.383

C. CAPM bond

Winning model **Losing model**			**Cb** **F5**	**Cb** **F7**
	Cb_Csb	Cb_E4	Cb_F5	Cb_F7
Sum of coeff. Diff.	0.796	-1.358	3.682**	4.797***
Wald p-Value	0.672	0.409	0.030	0.004
% of coeff. Diff. >0	46.670	37.78*	68.89***	80.00***
Binomial p-Value	0.383	0.07	0.01	0.00

D. CAPM stock + bond

Winning model **Losing model**		**Csb** **F5**	**Csb** **F7**
	Csb_E4	Csb_F5	Csb_F7
Sum of coeff. Diff.	-1.545	5.137**	6.230***
Wald p-Value	0.472	0.014	0.002
% of coeff. Diff. >0	42.220	73.33***	86.67***
Binomial p-Value	0.186	0.00	0.00

E. E4

Winning model **Losing model**	**E4** **F5**	**E4** **F7**
	E4_F5	E4_F7
Sum of coeff. Diff.	9.969***	10.45***
Wald p-Value	0.000	0.00
% of coeff. Diff. >0	82.22***	86.67***
Binomial p-Value	0.00	0.00

F. F5

Winning model **Losing model**	
	F5_F7
Sum of coeff. Diff.	4.420
Wald p-Value	0.345
% of coeff. Diff. >0	60.000
Binomial p-Value	0.116

Table 9: Model Horse Race – Sharpe Ratio Based on a GARCH Model

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using full bond fund sample. We estimate the relation between flow and a fund's decile ranking based on different performance measures by running the following regression:

$$F_{pt} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{pt-1} + \mu_t + e_{pt},$$

$F_{pt}$is the fund flow of mutual fund p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund p in month $t-1$ is in the decile i based on the first measure and decile j based on the second measure. We employ GARCH (1,1) model using 60-month past returns to estimate fund's variance to align with our method to estimate alphas. To estimate the model, the dummy variable for $i = 5$ and $j = 5$ is excluded. The matrix $X_{p,t-1}$ represents control variables including the lagged fund flow from month $t-1$, the lagged fund expense ratio, a dummy for no-load funds, a fund's return standard deviation estimated over the prior twelve months, the log of fund size and the log of fund age in month $t-1$ as well as Morningstar rating dummies in month $t-1$. We also include time fixed effects (μ_t).

For each pairwise comparison of two performance measures, we have 45 such b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji} to determine whether investors are more sensitive to the first or to the second measure (alpha estimated using the first model or using the second model), we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. If investors are more responsive to the first measure than the second one, we would expect to reject the null hypothesis in favor of the alternative hypothesis that b_{ij} is greater than b_{ji}. The table reports the results of two hypothesis tests for each horse race: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of difference equals 50%. We test the first hypothesis with a Wald test and the second one with a Binomial test. We present a model as "winning model" if the sum of coefficient differences is statistically significant from zero. A "–" indicates no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Monthly Sharpe Ratio Based on GARCH (1,1) Model

Winning model **Losing model**	**SR_GARCH** **RR**	**SR_GARCH** **Cb**	**SR_GARCH** **Csb**	**SR_GARCH** **E4**	**SR_GARCH** **F5**	**SR_GARCH** **F7**
	SR_GARCH_RR	SR_GARCH_Cb	SR_GARCH_Csb	SR_GARCH_E4	SR_GARCH_F5	SR_GARCH_F7
Sum of coeff. Diff.	13.73***	11.52***	6.726***	4.469**	7.550***	7.280***
Wald p-Value	0.000	0.000	0.004	0.022	0.000	0.000
% of coeff. Diff. >0	75.56***	77.78***	73.33***	62.22*	77.78***	80.00***
Binomial p-Value	0.000	0.000	0.001	0.068	0.000	0.000

Table 10: Model Horse Race – Alternative Factor Models

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using full bond fund sample. We estimate the relation between flow and a fund's decile ranking based on different performance measures by running the following regression:

$$F_{pt} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{pt-1} + \mu_t + e_{pt},$$

F_{pt} is the fund flow of mutual fund p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund p in month $t-1$ is in the decile i based on the first model and decile j based on the second model. To estimate the model, the dummy variable for $i = 5$ and $j = 5$ is excluded. The matrix $X_{p,t-1}$ represents control variables including the lagged fund flow from month $t-1$, the lagged fund expense ratio, a dummy for no-load funds, a fund's return standard deviation estimated over the prior twelve months, the log of fund size and the log of fund age in month $t-1$ as well as Morningstar rating dummies in month $t-1$. We also include time fixed effects (μ_t).

For each pairwise comparison of two performance measures, we have 45 such b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji} to determine whether investors are more sensitive to the first or to the second measure (alpha estimated using the first model or using the second model), we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. If investors are more responsive to the first measure than the second one, we would expect to reject the null hypothesis in favor of the alternative hypothesis that b_{ij} is greater than b_{ji}. The table reports the results of two hypothesis tests for each horse race: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of difference equals 50%. We test the first hypothesis with a Wald test and the second one with a Binomial test. We present a model as "winning model" if the sum of coefficient differences is statistically significant from zero. A "–" indicates no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 10: Model Horse Race – Alternative Factor Models (continued)

A. Augmented FF5 with bond MOM

Winning model	**MOMb**	**SR**	–	–	**E4**	–	–
Losing model	**RR**	**MOMb**	–	–	**MOMb**	–	–
	RR_MOMb	SR_MOMb	Cb_MOMb	Csb_MOMb	E4_MOMb	F5_MOMb	F7_MOMb
Sum of coeff. Diff.	-5.163**	7.670***	0.094	0.687	6.844*	-2.33	-2.631
Wald p-Value	0.033	0.000	0.9727	0.8191	0.061	0.6654	0.5691
% of coeff. Diff. >0	28.89***	71.11***	44.44	53.33	75.56***	37.78*	42.22
Binomial p-Value	0.003	0.003	0.2757	0.383	0.000	0.068	0.1856

B. Bai et al. (2018) factor model

Winning model	–	**SR**	–	–	**E4**	–	–
Losing model	–	**B4**	–	–	**B4**	–	–
	RR_B4	SR_B4	Cb_B4	Csb_B4	E4_B4	F5_B4	F7_B4
Sum of coeff. Diff.	1.707	15.13***	0.233	3.957	5.155*	0.356	1.018
Wald p-Value	0.6821	0.00	0.9495	0.1407	0.051	0.8886	0.6704
% of coeff. Diff. >0	55.56	91.11***	60	57.78	60	51.11	62.22*
Binomial p-Value	0.2757	0.00	0.1163	0.1856	0.1163	0.5	0.07

C. Chung et al. (2019) factor model

Winning model	–	**SR**	–	**Csb**	**E4**	–	–
Losing model	–	**C7**	–	**C7**	**C7**	–	–
	RR_C7	SR_C7	Cb_C7	Csb_C7	E4_C7	F5_C7	F7_C7
Sum of coeff. Diff.	-0.19	10.26***	1.974	3.956*	8.779***	2.99	0.916
Wald p-Value	0.923	0.000	0.3668	0.095	0.001	0.5462	0.82
% of coeff. Diff. >0	53.33	91.11***	66.67**	71.11***	86.67***	62.22*	62.22*
Binomial p-Value	0.383	0.000	0.018	0.003	0.000	0.068	0.068

D. Ludvigson-Ng macro factors

Winning model	**RR**	**SR**	**Cb**	**Csb**	**E4**	**F5**	**F7**
Losing model	**macro**	**macro**	**macro**	**macro**	**macro**	**macro**	**macro**
	RR_macro	SR_macro	Cb_macro	Csb_macro	E4_macro	F5_macro	F7_macro
Sum of coeff. Diff.	10.22***	17.84***	9.561***	9.556***	10.69***	8.000***	7.414***
Wald p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
% of coeff. Diff. >0	84.44***	95.56***	84.44***	88.89***	91.11***	88.89***	88.89***
Binomial p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000

E. Augmented Fama-French factor model (2015) with TERM and DEF

Winning model	–	**SR**	–	–	**E4**	–	–
Losing model	–	**FF7e**	–	–	**FF7e**	–	–
	RR_FF7e	SR_FF7e	Cb_FF7e	Csb_FF7e	E4_FF7e	F5_FF7e	F7_FF7e
Sum of coeff. Diff.	-1.283	9.321***	0.993	3.617	7.238***	2.68	2.802
Wald p-Value	0.5179	0.000	0.6386	0.1328	0.005	0.4538	0.387
% of coeff. Diff. >0	40	86.67***	57.78	62.22*	84.44***	68.89***	55.56
Binomial p-Value	0.1163	0.000	0.1856	0.068	0.000	0.008	0.2757

F. Augmented Hou et al. factor model (2015) with TERM and DEF

Winning model	–	**SR**	–	–	**E4**	–	–
Losing model	–	**HXZ**	–	–	**HXZ**	–	–
	RR_HXZ	SR_HXZ	Cb_HXZ	Csb_HXZ	E4_HXZ	F5_HXZ	F7_HXZ
Sum of coeff. Diff.	-2.843	8.606***	0.135	1.625	6.774**	0.393	-1.215
Wald p-Value	0.1525	0.000	0.9499	0.5174	0.011	0.9158	0.6841
% of coeff. Diff. >0	37.78*	82.22***	57.78	64.44**	75.56***	42.22	40
Binomial p-Value	0.068	0.000	0.1856	0.036	0.000	0.1856	0.1163

G. Augmented Stambaugh-Yuan factor model (2017) with TERM and DEF

Winning model	–	**SR**	–	–	**E4**	–	–
Losing model	–	**M4**	–	–	**M4**	–	–
	RR_M4	SR_M4	Cb_M4	Csb_M4	E4_M4	F5_M4	F7_M4
Sum of coeff. Diff.	-1.185	9.676***	0.552	2.957	7.747***	1.307	0.212
Wald p-Value	0.5491	0.000	0.7982	0.23	0.006	0.7609	0.9514
% of coeff. Diff. >0	40	91.11***	55.56	68.89***	84.44***	48.89	51.11
Binomial p-Value	0.1163	0.000	0.2757	0.008	0.000	0.5	0.5

Table 11: Controlling for Morningstar fixed income style box

This table presents the results of pairwise comparisons of different performance measures to explain fund flows using the full corporate bond fund sample. We estimate the relation between flow and a fund's decile ranking based on different performance measures by running the following regression:

$$F_{pt} = a + \sum_i \sum_j b_{ij} D_{ij,p,t-1} + cX_{pt-1} + \mu_t + e_{pt},$$

F_{pt} is the fund flow of mutual fund p in month t. $D_{ij,p,t-1}$ is a dummy variable that takes on a value of one if fund p in month $t-1$ is in the decile i based on the first model and decile j based on the second model. To estimate the model, the dummy variable for $i = 5$ and $j = 5$ is excluded. The matrix $X_{p,t-1}$ represents control variables including the lagged fund flow from month $t-1$, the lagged fund expense ratio, a dummy for no-load funds, a fund's return standard deviation estimated over the prior twelve months, the log of fund size and the log of fund age in month $t-1$, Morningstar rating dummies in month $t-1$ as well as Morningstar fixed income style box dummies in month $t-1$. We also include time fixed effects (μ_t).

For each pairwise comparison of two performance measures, we have 45 such b coefficient comparisons. With each pair of coefficients b_{ij} and b_{ji} to determine whether investors are more sensitive to the first or to the second measure (alpha estimated using the first model or using the second model), we test the null hypothesis that $b_{ij} = b_{ji}$ for all $i \neq j$. If investors are more responsive to the first measure than the second one, we would expect to reject the null hypothesis in favor of the alternative hypothesis that b_{ij} is greater than b_{ji}. The table reports the results of two hypothesis tests for each horse race: (1) H_0: The summed difference across all 45 comparisons is equal to zero, (2) H_0: The proportion of difference equals 50%. We test the first hypothesis with a Wald test and the second one with a Binomial test. We present a model as "winning model" if the sum of coefficient differences is statistically significant from zero. A "–" indicates no significant difference. The standard errors are double-clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 11: Controlling for Morningstar fixed income style box (continued)

A. Sharpe ratio

Winning model	**SR**	**SR**	**SR**	**SR**	**SR**	**SR**
Losing model	**RR**	**Cb**	**Csb**	**E4**	**F5**	**F7**
	SR_RR	SR_Cb	SR_Csb	SR_E4	SR_F5	SR_F7
Sum of coeff. Diff.	17.28***	13.90***	9.026***	6.338***	9.450***	9.539***
Wald p-Value	0.000	0.000	0.000	0.001	0.000	0.000
% of coeff. Diff. >0	91.11***	91.11***	84.44***	80.00***	82.22***	84.44***
Binomial p-Value	0.000	0.000	0.000	0.000	0.000	0.000

B. Raw return

Winning model	–	**Csb**	**E4**	–	–
Losing model	–	**RR**	**RR**	–	–
	RR_Cb	RR_Csb	RR_E4	RR_F5	RR_F7
Sum of coeff. Diff.	-5.133	-4.295*	-5.231**	-1.099	-0.528
Wald p-Value	0.1624	0.081	0.020	0.605	0.7991
% of coeff. Diff. >0	33.33**	28.89***	17.78***	44.44	42.22
Binomial p-Value	0.018	0.003	0.000	0.2757	0.1856

C. CAPM bond

Winning model	–	**E4**	–	–
Losing model	–	**Cb**	–	–
	Cb_Csb	Cb_E4	Cb_F5	Cb_F7
Sum of coeff. Diff.	-2.569	-4.083*	0.602	1.406
Wald p-Value	0.323	0.084	0.796	0.537
% of coeff. Diff. >0	40.000	28.89***	60.000	60.000
Binomial p-Value	0.116	0.003	0.116	0.116

D. CAPM stock + bond

Winning model	–	**Csb**	**Csb**
Losing model	–	**F5**	**F7**
	Csb_E4	Csb_F5	Csb_F7
Sum of coeff. Diff.	-3.815	4.116*	4.738**
Wald p-Value	0.169	0.093	0.048
% of coeff. Diff. >0	40.000	71.11***	75.56***
Binomial p-Value	0.116	0.003	0.000

E. E4

Winning model	**E4**	**E4**
Losing model	**F5**	**F7**
	E4_F5	E4_F7
Sum of coeff. Diff.	8.315***	8.283***
Wald p-Value	0.001	0.000
% of coeff. Diff. >0	80.00***	84.44***
Binomial p-Value	0.000	0.000

F. F5

Winning model	–
Losing model	–
	F5_F7
Sum of coeff. Diff.	4.945
Wald p-Value	0.325
% of coeff. Diff. >0	55.560
Binomial p-Value	0.276

Table 12: Univariate Flow-Performance Sensitivity Estimations

This table reports the beta estimates from the following equation for different risk models:

$$\beta_{flow,performance} = \frac{cov(\Phi(flow_{i,t-1}), (\Phi(\alpha_{i,t-1}))}{var(\Phi(\alpha_{i,t-1}))} > 0,$$

where Φ is a function that returns the sign of a real number, taking values of 1 for a positive number, -1 for a negative number and 0 for zero. The sample period is from 1996 to June 2017. For the ease of interpretation, the table reports $(\beta_{flow,performance} + 1)/2$ which denotes the average probability that the sign of the fund flow [*S(flow)*] is positive (negative) conditional on the sign of the performance measure [*S(performance)*] being positive (negative). Each row corresponds to a different performance measure. p-Values are based on a *t*-test of $\beta_{flow,performance}$ using double-clustered standard errors (by fund and month).

	S (Flow)	p-Value
S (Sharpe ratio)	54.25	0.00
S (Raw return)	54.18	0.00
S (CAPM bond alpha)	53.02	0.00
S (CAPM bond and stock alpha)	53.22	0.00
S (E4 alpha)	52.78	0.00
S (FF5 bond alpha)	52.84	0.00
S (F7 alpha)	52.83	0.00

Table 13: Flow-Performance Model Horse Race: Berk and Van Binsbergen (2016) Pairwise Model Comparisons

This table reports the results from pairwise comparisons of raw returns, Sharpe ratio and different alphas as in Berk and Van Binsbergen (2016). The first two columns provide the coefficient estimate and double-clustered (by fund and month) *t*-statistic of the univariate regression of signed flows on signed out-performance. The rest of the columns provide the statistical significance of the pairwise test coefficient b_1 in the following equation:

$$\Phi(Flow_{it}) = a + b_1 \left(\frac{\Phi(\alpha^{m1}_{it-1})}{var(\Phi(\alpha^{m1}_{it-1}))} - \frac{\Phi(\alpha^{m2}_{it-1})}{var(\Phi(\alpha^{m2}_{it-1}))} \right) + \xi_{it},$$

where we compare the flow-performance regression coefficients, $\beta_{flow,performance}$ of two models m1 and m2.

	β	Uni. t-stat	Sharpe ratio	Return	CAPMbond	CAPMstkb	E4	F5	F7
Sharpe ratio	0.0850	8.98	0.00	-0.37	2.11	2.07	4.67	3.95	4.05
Return	0.0836	8.58		0.00	1.86	2.26	4.58	3.97	4.06
CAPMbond	0.0603	7.62			0.00	0.95	4.81	3.13	3.13
CAPMstkb	0.0643	8.92				0.00	4.05	2.59	2.68
E4	0.0556	9.98					0.00	-1.18	-1.14
F5	0.0568	7.80						0.00	0.09
F7	0.0567	7.93							0.00