



HAL
open science

Tweeting for Money: Social Media and Mutual Fund Flows

Juan Imbet, Javier Gil-Bazo

► **To cite this version:**

Juan Imbet, Javier Gil-Bazo. Tweeting for Money: Social Media and Mutual Fund Flows. 2024. hal-04726330

HAL Id: hal-04726330

<https://hal.science/hal-04726330v1>

Preprint submitted on 8 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Tweeting for Money: Social Media and Mutual Fund Flows*

Javier Gil-Bazo[†]

Juan F. Imbet[‡]

September 6, 2024

Abstract

We unveil asset managers' social media communications as a distinct new channel for attracting flows of money to mutual funds. Combining a database of more than 1.6 million posts on X/Twitter by U.S. mutual fund families with textual analysis, we find that flows of money to mutual funds respond positively to both the number and tone of the posts. The link between social media communications and flows of money is not explained by marketing efforts, but the two strategies reinforce each other. A high-frequency analysis that exploits intraday ETF trade data allows us to isolate the effect of tweets on investor decisions from potential confounders. We then consider and test four different economic mechanisms. The results of these tests do not support the hypothesis that asset managers' social media communications reduce search costs for potential investors. The results do not support, either, that asset management companies' Twitter activity increases investor attention or alleviates information asymmetries by communicating performance-relevant information to investors. In contrast, our evidence suggests that asset managers use social media as an effective persuasion tool.

Keywords: social media; Twitter; mutual funds; machine learning; textual analysis; search costs; information asymmetry; information frictions; persuasion.

JEL Classification: G11; G23; D83

*The authors thank Simona Abis, Vicente Bermejo, Patrick Bolton, Tony Cookson, Victor DeMiguel, Ariadna Dumitrescu, Rich Evans, Javier Gomez Biscarri, Philip Hamill, Filippo Ippolito, Alberto Manconi, Andrea Polo, Fabrice Riva, Emanuele Rizzo, Maria Tselika and Fernando Zapatero as well as seminar participants at CUNEF, Universitat Pompeu Fabra, Université Paris Dauphine, Kozminski University and conference participants at the 2021 International Conference of the French Finance Association, the 2021 Finance Forum of the Spanish Finance Association, the 2022 FMA Consortium on Asset Management and Fintech, the 2022 British Accounting and Finance Association SWAG Annual Conference and the 2023 Midwestern Finance Association for helpful comments and suggestions. Javier Gil-Bazo acknowledges financial support from the Spanish Ministry of Science and Innovation (research grant PID2020-118541GB-I00), and Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S). Both authors acknowledge financial support from the FinTech & Digital Finance Chair at Dauphine-PSL Paris.

[†]Universitat Pompeu Fabra, Barcelona School of Economics, and UPF Barcelona School of Management, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain, e-mail:javier.gil-bazo@upf.edu

[‡]Corresponding author. Université Paris Dauphine and PSL Research University, Pl. du Maréchal de Lattre de Tassigny, 75016 Paris, e-mail: juan.imbet@dauphine.psl.eu

1 INTRODUCTION

As of June 2024, more than 5 billion people in the world, 84% of the population aged 18 years and older, were using social media to communicate with others, entertain themselves, and stay informed.¹ Social media has the ability to inform the public by providing instant news updates, facilitating the sharing of ideas, and connecting individuals. However, the same features that make social media a powerful tool for information sharing can also be exploited for influencing individuals' beliefs and their actions. In the context of financial markets, social media provides companies with a powerful tool to communicate with investors and alleviate information frictions by disclosing relevant information in real time with few constraints on timing or content. However, there are growing concerns that social media can be used to influence investors' views and even to manipulate prices.² In this paper, we study social media communications in a financial market characterized by information frictions and populated by unsophisticated investors, the market for mutual funds. More specifically, we study the impact of asset management companies' use of social media (X/Twitter) on mutual fund investor decisions. Our findings unveil social media as a novel channel for attracting flows of new money to mutual funds, which is distinct from fund families' marketing efforts. Our tests do not support several hypotheses for the economic mechanism underlying the link between social media activity based on theories of search costs, attention, and information asymmetries, and point, instead, towards persuasion.

Social media allows asset management firms to communicate with current and prospective investors without the strict constraints imposed by mandatory information disclosures on the timing, content, and framing of information.³ Asset management companies have incentives to use the flexibility of social media to alleviate information frictions, such as search costs and information asymmetries that plague the market for mutual funds (Sirri and Tufano, 1998; Hortaçsu and Syverson, 2004; Roussanov et al., 2021). For instance, asset management companies may use social media to help investors locate key information, such as fees and performance that are part of mandatory disclosures, but not easy to find by retail

¹Data from <https://datareportal.com/social-media-users>.

²SEC (2018), "Elon Musk Charged With Securities Fraud for Misleading Tweets," (<https://www.sec.gov/news/press-release/2018-219>); Spichak (2021), "Elon Musk Hopes SEC Will Investigate Him over Dogecoin Tweets: 'It Would Be Awesome'," Newsweek, (<https://www.newsweek.com/elon-musk-sec-investigation-dogecoin-bitcoin-cryptocurrency-tweets-1572290>) Mohamed (2021), "Big Short' investor Michael Burry says he'll stop tweeting after SEC regulators paid him a visit," Businessinsider.com; (<https://markets.businessinsider.com/currencies/news/big-short-investor-michael-burry-stop-tweets-sec-regulators-visit-2021-3-1030222890>).

³Advertisement and retail investor communication by asset management companies must comply with SEC rule 482 and FINRA rule 2210. In 2003, SEC rule 482 modified the Securities Act of 1933-Section 5 that stated that all fund advertisement must have information that is contained in the statutory prospectus. With rule 482, investment companies are allowed to include information not included in the statutory prospectus. This allows investment companies to include up-to-date information in rule-482 advertisements, such as information about current economic conditions that are not commonly included in a fund's prospectus. FINRA Rule 2210 governs communications with the public including communications with retail and institutional investors. The rule provides standards for the content, approval, record-keeping and filing of communications with FINRA. The rule prohibits false, exaggerated, unwarranted, and misleading information communications, as well as projections of future performance.

investors. Companies can also use social media to bring investor attention towards their funds. Finally, social media is an ideal vehicle for conveying soft information that is potentially relevant for future fund performance. For instance, they may convey asset managers' views about the markets in which they invest.

On the other hand, asset management companies also have incentives to *persuade* investors. [Mullainathan and Shleifer \(2005\)](#) and [Mullainathan et al. \(2008\)](#) show theoretically how persuaders can take advantage of individuals' associative thinking to create *uninformative* messages that associate their product with a given category and its attributes. For example, asset management companies may want to associate themselves with enhancers of opportunities in times of high market returns and with trustworthy advisers in down markets. Note, however, that the ability of asset managers to influence investors' opinions through social media is limited by the fact that in social media, unlike in traditional media, communication is bidirectional. Users are not passive receivers of messages. Instead, they can reply to the posts of firms in a critical way. Such interactions are observable by other users, who can amplify messages by *liking*, reposting or posting them across different platforms in real time and any time of the day. These features potentially limit the extent to which managers can influence investor perception of their products ([Cade, 2018](#)).

The mutual fund industry is an ideal laboratory to study the use of social media in financial markets, as thousands of actively managed mutual funds compete for investors' money in a context of search costs and information asymmetry.

Twitter is also particularly appropriate for our purposes given its rising influence among investors. A fast-growing literature provides evidence that Twitter activity can predict prices of stocks and other asset classes ([Bollen et al., 2011](#); [Ranco et al., 2015](#); [You et al., 2017](#); [Gholampour and van Wincoop, 2019](#); [Gu and Kurov, 2020](#) [Cookson et al., 2023](#)). Importantly, some recent studies show how communications on Twitter not just *predict* future returns but also *impact* asset prices ([Bianchi et al., 2023](#); [Bianchi et al., 2024](#); [Cookson et al., 2023](#)). A recent work by [Cookson et al. \(2024b\)](#) surveys the fast growing literature on the role of social media in Finance. Consistently with the increasing relevance of social media in financial markets, the presence of asset management firms on Twitter has grown at a very fast pace in the last years. In our sample, the number of posts on Twitter (tweets) by all mutual fund families went from almost zero prior to 2009 to around 10,000 tweets per month in 2020.

To investigate whether and how asset management companies influence investors' decisions through social media communications, we build a database containing more than 1.6 million Twitter posts by mutual fund families managing domestic equity funds in the U.S. between January 2009 and December 2020. We then use these data to measure the intensity of Twitter activity by asset management firms at

the monthly frequency, as well as the tone of their tweets (using the VADER Sentiment Analysis tool). The tone score from VADER allows us to classify tweets as positive, neutral, and negative. In addition, we use an unsupervised topic modelling algorithm to study the content of those tweets. This analysis suggests that tweets by asset management firms fall in one of three main categories. Market Commentary, Customer Service, and Financial Advice. Finally, we merge the Twitter data with data on U.S. domestic equity mutual funds obtained from the CRSP Survivor-Bias-Free U.S. Mutual Fund database. In our sample, 269 out of 848 firms managing U.S. equity funds have a Twitter account and post at least one tweet during the sample period.

Our findings can be summarized as follows. First, families that use Twitter tend to manage more assets, more funds, funds in more investment categories, and have better recent performance than families that do not use Twitter, which suggests that economies of scale play a role in the decision to implement a social media strategy. Among those firms that use Twitter, intensity of usage is higher among younger firms, firms managing less risky funds, and firms with better overall performance. Across all non-index U.S. diversified equity funds, fund families that use Twitter collectively manage 3.16 USD trillion in an average month, which corresponds to 65.2% of total assets. Moreover, they operate 6,252 funds in an average month, 52% of the total (12,031).

Second, the number of tweets posted by a fund family in a given month predicts significantly higher flows to the family's funds in the following month. The increase in flows is economically significant. A one standard deviation increase in the number of tweets from the unconditional mean is associated with extra flows to the family's funds of 0.035% per month. Since the average family in the Twitter sample manages assets worth USD 21,799.40 million, this growth in assets amounts to USD 7.657 million. This result is robust to different ways of modelling the flow-performance relationship, to the inclusion of time, fund, and fund family fixed effects, and to sample selection bias correction.

Third, we also find that the tone of tweets matters. While the majority of tweets have a neutral tone, asset management companies also post tweets that are classified as positive or negative. While a larger number of positive tweets is followed by larger net flows of money, non-positive tweets in a given month predict lower net flows. This link between the tone of posts and flows is not explained by the tone capturing different messages about different topics. Exploiting variation in tone across tweets with the same topic, we continue to find that more positive tweets predict higher flows, although this relation is not statistically significant for Financial Advice tweets. These findings suggest that fund families' tweets do not influence flows by simply providing customer service.

The possibility to attract investors' money by posting positive tweets raises the question of whether fund families face any trade-off when posting positively. To address this question we construct measures

of consistency between the tone of the tweets and subsequent fund performance. Our results suggest that investors seem to value this consistency. Also, they seem to penalize funds of families posting positive tweets if the subsequent performance of those funds is poor.

Fourth, exploiting data on users' reactions to tweets ("likes", retweets, and replies), we find that, conditional on the number of tweets posted by a fund family, a larger number of interactions with those tweets further increases flows.

Fifth, the link between Twitter activity and mutual fund flows is not explained by this activity being correlated with asset management companies' marketing efforts, although the two strategies appear to reinforce each other. Our results are not explained, either, by fund families using social media to reiterate information on a number of known events that may influence flows, such as a change in the fund's management, the presence of funds in the family with very recent stellar performance, or any news about the fund family disseminated through social media by third parties.

Sixth, to identify the effect of asset management companies' tweets on investor decisions, we combine a high-frequency approach with intraday ETF trading data. The analysis is based on the idea that asset management companies' positive tweets increase the demand not just for their open-end mutual funds but also for their ETFs. Since ETF shares are traded in an exchange, we expect a family's positive tweets to put upward pressure on the price of its ETFs. More specifically, we study whether an ETF's price changes in a very short time window around the time when the ETF's family tweets. This high-frequency approach allows for clean identification of the effect of tweets on prices with the identifying assumption being that over such a short window of time no other relevant information affecting the security's price is released (Bianchi et al., 2023, 2024). Our results indicate that a positive tweet by an asset management company increases the price of its ETFs by 0.65 basis points relative to other tweets with a neutral tone in the first 35 minutes after the tweet is posted. Consistently with the effect on ETF prices being driven by increased demand for the family's products and not fundamentals about the underlying assets, the price increase reverts over a longer period. These findings provide support for the hypothesis that asset management companies' tweets affect fund investor decisions through social media.

Finally, we consider and test four possible mechanisms for the impact of fund families' tweets on mutual fund flows. We first test the hypothesis that asset management companies use social media to reduce search costs for investors. According to this hypothesis, tweets should increase in net flows but would not affect outflows, since existing investors have already paid the search cost of the fund (Sirri and Tufano, 1998; Hortaçsu and Syverson, 2004). Also, a reduction in search costs should decrease the convexity of the relation between inflows and performance, particularly among the best performing funds, since new investors would be willing to consider a larger set of funds and not just recent winners (Huang et al., 2007).

To test these implications, we analyze how the number of tweets affects inflows and outflows of money and their relation to fund performance. Our evidence does not support the theory’s predictions. Tweets do not just increase inflows but they also decrease outflows. Moreover, the relation between inflows and performance does not become flatter.

The results of this analysis do not support an attention hypothesis, either. Models of rational inattention predict that increases in investor attention to stocks result in more retail investors considering both purchases and sales of stocks, although the increase in purchases by retail investors should be higher than the increase in sales due to costs and constraints affecting short sales (Barber and Odean, 2008; Da et al., 2011). The increase in inflows following a higher number of tweets that we observe in the data is consistent with this story. However, we also find that a higher number of tweets *reduces* outflows across funds, which contradicts the theoretical prediction. Moreover, we also find that outflows become *less* sensitive to performance in the low performance region, which is at odds with the notion of attentive investors discriminating among the family’s funds.

We then investigate whether asset management companies use social media to reduce information asymmetries, i.e., to convey to investors information that predicts future fund performance and is not available to the public. We find no evidence that the number or tone of tweets of an asset management company predicts superior future fund performance, controlling for fund characteristics, and particularly for fund size and flows to account for diseconomies of scale, and using different measures of fund performance.

Finally, we test a theoretical prediction of the persuasion hypothesis by (Mullainathan and Shleifer, 2005; Mullainathan et al., 2008). In particular, persuasion requires that the message conveyed to investors be adapted to different circumstances. In times of low investor sentiment, asset managers will want to associate their products with safety and trust, whereas in times of high investor sentiment, asset managers will favor the idea of seizing investment opportunities. Consequently, successful persuasion will lead investors towards safer products such as money market or bond funds in low-sentiment markets and towards riskier products, such as equity funds, in high-sentiment markets. Since our sample consists of only equity funds, we would expect the Number of Tweets and Number of Positive Tweets to be more strongly associated with flows in times of high investor sentiment. Moreover, we would expect retail investors, who are less sophisticated than institutional investors and more vulnerable to persuasion, to be more sensitive to tweets in times of high sentiment. Our results confirm these predictions.

Roussanov et al. (2021) show that most funds are too big relative to the efficient size implied by the model of Berk and Green (2004) and deliver substantially negative abnormal returns to investors, while a few funds with the most skilled managers are too small and thus are able to outperform, which is con-

sistent with the presence of information frictions. While social media communications can potentially alleviate information frictions and increase the allocative efficiency of the market, our results point towards social media being used to increase asset managers' fee revenues while providing little benefit to investors.

We contribute to the literature by unveiling asset management companies' social media communications as a new and distinct channel for influencing mutual fund investors' decisions. Our paper is related to studies that investigate advertising in the mutual fund industry. [Sirri and Tufano \(1998\)](#) show that marketing effort, as proxied by fund fees, is associated with larger fund flows. [Jain and Wu \(2000\)](#) find that management companies tend to advertise funds with higher recent performance. [Cronqvist \(2006\)](#) investigates the content of mutual fund advertisements in Sweden and finds that most fund ads are not informative about fund quality and yet influence individuals' portfolio decisions. [Gallaher et al. \(2015\)](#) show that mutual fund families' advertising expenditures attract flows to the family's funds as well as to other funds in the industry, reduce redemptions, and increase the convexity of the flow-performance relationship. Our paper is also related to the study of [Hillert et al. \(2023\)](#), who find a positive association between the tone of the text in shareholder letters from asset management companies and subsequent fund flows.

Social media is fundamentally different from advertising or shareholder letters in that it is bidirectional. Users can interact with firms' messages and counteract or amplify their effect. Our findings confirm the importance of marketing in delegated portfolio management, but also show that social media communications are distinct from marketing. Their effect on flows is independent of that of marketing expenses. More importantly, unlike studies of mutual fund marketing, we find that social media communications reduce search costs for investors. Therefore, unlike advertising or marketing, the role of social media is not limited to persuasion. Another important distinction is that social media communications take place in real time, which allows us to determine with full precision the time at which messages are released. In [section 6](#) we exploit this feature to isolate the effect of Twitter posts on ETF share prices from potential confounders.

Our paper also contributes more generally to the literature on non-mandatory corporate disclosures ([Kim and Verrecchia, 1991](#); [Dye and Sridhar, 2004](#); [Cornelli et al., 2013](#); [Bertomeu and Marinovic, 2016](#)), and to the recent literature of textual analysis in Finance and Accounting (see [Loughran and McDonald, 2016](#)).⁴ The study of [Blankespoor et al. \(2014\)](#) is particularly relevant to our paper. These authors show that when public firms use Twitter to disseminate firm-initiated news, information asymmetries decline, as evidenced by narrower bid-ask spreads. In contrast, our results suggest that Twitter does not help

⁴Recent studies have applied machine learning to study fund's behavior including analyzing their prospectuses ([Abis and Lines, 2022](#)), or their corporate governance preferences ([Bubb and Catan, 2022](#)).

alleviate information asymmetries in the mutual fund industry.

2 DATA

In this section, we describe the data used in the analysis. We draw mainly on two data sets, the CRSP Survivor-Bias-Free U.S. Mutual Fund database and a database of tweets from January 2009 to December 2020 posted by mutual fund families. From the former, we obtain information at the share class level on returns, assets under management, investment category, expenses, ticker, manager information, and age. Even though our Twitter database starts in 2009, we collect mutual fund data from 2006 so we can use three years of prior historical data to estimate risk-adjusted returns.

For our main analysis, we keep only non index U.S. diversified equity funds following CRSP style level one ($USE = 1$ and $index_fund_flag = 0$). To avoid discrepancies between the objective reported by CRSP and the Lipper classification of the fund we manually discard those funds whose Lipper objective class name does not correspond to an equity fund. Among these funds we discard less than one percent of the funds that are classified as municipal debt funds, money market funds, debt funds, and funds focusing on other securities.

To construct variables at the mutual fund level, we follow the same share aggregation procedure as [Gil-Bazo and Ruiz-Verdú \(2009\)](#). Total Net Assets (TNA) of a fund are the sum of the TNA of the fund's share classes. Returns and expense ratios are TNA-weighted averages across all share classes in the fund. The age of the fund is the age of the oldest share class in the fund. Following [Berk and van Binsbergen \(2016\)](#) and, [Pastor et al. \(2015\)](#) we exclude observations for funds with less than USD 15 million in TNA. Finally, we identify if a fund is institutional or retail using Morningstar data following [Dumitrescu et al. \(2022\)](#). In particular, we collect the institutional flag from Morningstar, which has been used in the mutual fund literature to identify retail and institutional funds by [Evans and Fahlenbrach \(2012\)](#) and [Leipold and Rueegg \(2020\)](#), among others. We then construct an indicator, Retail, for funds that have more than 50% of their assets under management in retail share classes.

To create some of our variables, we aggregate data at the fund family level based on the CRSP identifier `mgmt_code`. TNA at the family level is the sum of the TNA of each fund in the family, the age of the fund family is the age of the oldest fund in the family, and flows, expenses and returns are weighted averages across all funds in the family (based on the TNA of each fund in the family).⁵

For a subsample of funds, and until 2016, we obtain data on inflows and outflows, as in [Christoffersen et al. \(2013\)](#) and [Ha and Ko \(2019\)](#). These data are reported in SEC's N-SAR form, Item 28, which

⁵We compute flows at the family level using a weighted average of the flows of each fund in the family since we discard money market, bond, and hybrid funds from the sample before aggregation.

includes cash-flow information on a monthly basis at the portfolio level.⁶

Given the findings of Barber et al. (2016) and Berk and van Binsbergen (2016) that investors appear to use the CAPM to evaluate mutual fund performance, throughout the paper we focus on CAPM alphas as a determinant of flows, although in unreported results and in the predictability analysis we verify the robustness of our findings to using the three, four, five and six factor models of Fama and French (1993, 2015) and the momentum factor in Carhart (1997) to estimate performance.⁷ We compute the risk-adjusted return, $\hat{\alpha}_{it}$, of fund i in month t as the intercept plus the residual of the CAPM model:

$$\hat{\alpha}_{it} = r_{it}^e - \hat{\beta}_{it}r_{mt}^e, \quad (1)$$

where r_{it}^e is the excess return of fund i at month t over the risk free rate and r_{mt}^e is the excess return of the market portfolio over the risk free rate. We obtain the monthly risk-free rate and the market portfolio return from Prof. Kenneth French’s website and $\hat{\beta}_{it}$ is estimated for each fund and month t by running OLS rolling regressions of excess returns on market excess return over the three-year period ending in month $t - 1$. If less than three years of data are available in a given window, we require the fund to have at least 30 months of data and run the regressions with the data available.

To construct the database of mutual fund families’ Twitter accounts we obtain the names of all asset management companies in the CRSP database managing U.S. equity funds. Then, we perform a manual search through each one of the family names represented in the variable `mgmt_name` in the CRSP database and group similar names using the CRSP aggregation variable `mgmt_code`. Finally, we search for each family’s Twitter account in the asset management company’s website. In case that more than one Twitter account is used by the fund family, we use the tweets from all accounts.

Once the list of Twitter accounts is collected, we obtain all tweets from accounts that are active on three different dates: February 2018, December 2021, and July 2023. It is important to notice that if a fund family was on Twitter and decided to cancel its Twitter account before February 2018, we would not be able to get information on that family’s tweets. Similarly, for accounts that existed in February 2018 and were canceled before December 2021, we would have no information in that sub-period. To ensure we downloaded all information, we compare the earliest tweet obtained for each company with the true first tweet of the account as provided by Twitter.⁸ Our database contains 1,664,188 tweets from 284 different usernames, from January 2009 to December 2020.

⁶We thank Yeonjeong Ha and Kwangsoo Ko for kindly sharing their data with us.

⁷Evans and Sun (2020) show that mutual fund flows have become more sensitive to three-factor abnormal returns since Morningstar changed its methodology to compute fund ratings that account for funds’ investment style.

⁸The first tweet of any active account was found using the webpage <https://discover.twitter.com/first-tweet>, which is no longer available, although other websites provide the same service.

We score the sentiment of each tweet with Valence Aware Dictionary and sEntiment Reasoner (VADER), a sentiment classifier designed specifically for social media. [Hutto and Gilbert \(2014\)](#) propose VADER as a sentiment classifier for short messages that contain social text, like tweets, and show that it outperforms naive dictionary-based methods. VADER does not require training data. Instead, it relies on a human-validated sentiment lexicon and general grammatical rules, which makes it a transparent and easily adaptable method. It is particularly appropriate for texts that are commonly used in social media posts, which tend to employ a specific vocabulary and punctuation, as well as emoticons. Because of these advantages, VADER is becoming a popular sentiment analysis tool in the academic literature that studies social media in Finance ([Long et al., 2023](#); [Cookson and Niessner, 2023](#); [Cookson et al., 2023, 2024a](#)). Specifically, we use the VADER algorithm to compute a raw sentiment score based on the words contained in the tweet. The VADER algorithm is an aggregate of a negative component, a neutral component, and a positive component that load on negative, neutral and positive sentiment tokens, respectively. To compute sentiment, the VADER algorithm aggregates across these sub-components to form a sentiment score for each tweet, which is a number between -1 (very negative) and 1 (very positive). We classify a tweet as positive (negative) if the composite VADER score is greater (less) than 0.05 (-0.05), and classify it as neutral otherwise.

Figure 1 displays the total number of tweets across all mutual fund families, as well as the number of positive, negative, and neutral tweets per month. The figure shows a sharp increase in Twitter usage by mutual fund families over the sample period, with a peak in 2016. As expected, neutral tweets predominate, followed by positive and negative tweets.⁹

Out of 848 fund families in the final CRSP sample, 269 fund families tweet at least once during the sample period. This is the subsample we use in most of our analysis. Across all months in our sample, fund families that use Twitter collectively manage 3.16 USD trillion in an average month across their domestic equity funds, which corresponds to 65.2% of the total assets of non-index U.S. diversified equity funds. Moreover, they operate 6,252 different funds on average, which is 52% of the total number of funds in the sample. To understand how this subsample differs from the rest, Table 1 presents descriptive statistics of both fund and family characteristics for the Twitter subsample and the full sample. At the fund level, there are no clear differences between funds managed by fund families in the Twitter subsample and funds in the entire sample. However, at the family level, differences between fund families in both samples become more evident. Fund families in the Twitter subsample are on average older, manage more assets, more funds, and funds in more different investment categories.

⁹In unreported tests, we check the robustness of our results using a series of supervised machine learning algorithms with a training sample manually constructed, as well as using Large Language Models (LLMs) for sentiment analysis.

We explore the content of the tweets in our sample using a Latent Dirichlet Allocation (LDA) topic modelling algorithm (Blei et al., 2003). Under the LDA algorithm, tweets are represented as a mixture of topics, and a topic can be understood as a distribution of words (or lemmas). The LDA is a hierarchical Bayesian factor model for discrete data that has shown success in reducing the dimensionality of textual data in economic applications (Hansen et al., 2018; Bandiera et al., 2020). We estimate the LDA algorithm with three topics over the entire sample of tweets after removing common stop-words in tweets, lemmatizing the remaining words, and considering common bigrams in the sample. We identify that the three topics estimated correspond to three different types of communication: i) *Market Commentary*, ii) *Customer Service* (or general communications), and iii) *Financial Advice*. Figures 2, 3, and 4 present word-clouds of the three topics estimated in which the size of the word is proportional to its relevance in the topic. We label the first topic as *Market Commentary* as the words *global*, *ipo*, *investors*, *markets*, *investment*, *investing* and *equity* are the most common within the topic. We label the second topic as *Customer Service* as the words *help*, *us*, *hi*, *please*, *thanks*, *call*, *number* and *sorry* are the most common within the topic. Finally, we label the third topic as *Financial Advice* as the words *help*, *new*, *financial*, *bank* and *great* are the most common within the topic.

To save on notation we define the set of topics as

$$\mathcal{T} = \{\text{Market Commentary, Customer Service, Financial Advice}\} \quad (2)$$

The LDA algorithm provides weights (or coordinates) across these three topics such that each tweet can be represented with the tuple

$$\theta = (\theta^j)_{j \in \mathcal{T}} \quad (3)$$

such that $\theta^j \geq 0 \forall j \in \mathcal{T}$, $\sum_{j \in \mathcal{T}} \theta^j \leq 1$, and the remaining weight $1 - \sum_{j \in \mathcal{T}} \theta^j$ captures potential omitted topics not considered in the analysis, rounding errors in the estimation, and tweets that cannot be explained with the three topics. In our sample the average sum of weights equals 99.99 %. In unreported tests we estimate the model using two and four topics and find that in the first case the two topics resemble the market commentary and customer service identified in the paper, and in the second case the fourth topic has no clear separation from the other three topics. Moreover, we compare the *coherence score* of the trained models, a measure of out-of-sample performance, and conclude that three topics properly explain the distribution of tweets in our sample.

We define a tweet k as conveying information about topic l if not all of its coordinates are zero and

the weight on topic l is the largest.¹⁰

$$\tau_{kt}^l = \begin{cases} 1, & \text{if } l = \arg \max\{\theta^i\} \text{ and } \max\{\theta^i\} \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Using this definition, we also compute topic specific counts of tweets aggregating only over tweets classified with each of the three topics for positive, neutral and negative tweets. Our database contains 981,721 (59%) tweets classified as *Market Commentary*, 420,229 (25.25%) as *Customer Service*, and 262,238 (15.75%) as *Financial Advice*. Out of the *Market Commentary* tweets, 54.85 % are classified as positive, 13.38% as negative, and 31.75% as neutral. For *Customer Service*, and *Financial Advice* these numbers are (70.5%, 9.3%, 20.19%) and (57.83%, 8.3%, 33.87%) respectively.

We illustrate how our classification procedure works in practice by examining five tweets posted by asset management companies in our sample. The first example, presented in Figure 5, corresponds to a post written by Northern Trust which is classified as negative with a score of 0.994. The second tweet presented in Figure 6 written by PIMCO is classified as positive with a score of 1. Finally, we provide examples of the topic modelling of tweets. Figure 7 provides an example of a tweet classified as being related to *Market Commentary*, Figure 8 provides an example of a tweet classified as being related to *Customer Service*, and finally Figure 9 provides an example of a tweet classified as being related to *Financial Advice* together with their tone classification.

Finally, for part of our analysis, we obtain intraday data on ETF prices and volumes from FirstRate Data. Prices are adjusted for both splits and dividends. Using the ticker code of the ETF and focusing on ETFs within fund families that have a Twitter account, we are able to match intraday data for 867 different ETF tickers in our sample.

3 DETERMINANTS OF TWITTER ACTIVITY BY MUTUAL FUND FAMILIES

We start our analysis by investigating the determinants of Twitter usage by fund families. Although social media communication has low explicit costs, the implicit costs are non-trivial. Managing a social media communication strategy requires that social media managers coordinate with the marketing department and senior management in the process of setting goals and engaging with the public. In addition, social managers must create contents, promote the firm’s social media presence, and carefully monitor and evaluate the whole process. Since such costs are likely to have a fixed component, we expect larger asset

¹⁰We impose this restriction since around 0.3% of tweets cannot be explained by these three topics and have zero weights on all topics. These tweets tend to be short, use images, post urls or contain spelling mistakes which makes it improbable for the LDA algorithm to explain them.

management firms to be more likely to use social media. To proxy for family size, we use the log of the family’s total assets under management, the number of funds and the number of different categories (both in logs) in which families offer funds. Considering the potential benefits of social media, we conjecture that younger asset management firms have stronger incentives to use social media to gain visibility. Also, it seems plausible to think that asset management companies are more likely to communicate with investors when their funds have recently experienced better performance. We compute family age and performance as the age of the oldest fund in the family and the asset-weighted average CAPM alpha of the funds in the family over the previous 12 months, respectively. In addition to these variables of interest, we control for the characteristics of funds aggregated at the family level: asset-weighted average expense ratio; and asset-weighted average volatility of fund returns in the previous 12 months.

We analyze both the extensive and the intensive margins of fund families’ Twitter usage. We first estimate a linear probability model in a panel data at the family-month level with the Number of Tweets > 0 indicator as the dependent variable. In this regression, we use the full sample. Explanatory variables are lagged one month and we include time, and fund family fixed effects. Estimation results are presented in columns (1)-(3) of Table 2. As expected, all three proxies for family size are positively and significantly associated with the probability of tweeting. Moreover, family age is negatively associated with this probability, which indicates that younger families are more likely to use twitter at any point. Better fund performance and lower risk are both associated with a higher probability of using Twitter. In columns (2) and (3) we also find that fund families with lower expense ratios are more likely to use twitter in a given month.

We then regress the natural logarithm of 1 plus the Number of Tweets on the same set of explanatory variables as in the previous regression. In this case, we naturally restrict the sample to families that tweet at least once in our sample period. We include family and time fixed effects and compute standard errors clustered at the month level. Estimation results are presented in columns (4)-(6) of Table 2. Conditional on having a Twitter account and using it at least once in the sample, larger firms, as measured by total assets, tend to tweet more. However, families with funds in more investment categories tend to tweet less frequently. The sign on age remains negative but it is only significant in column (4) when using total assets as a proxy of size. Fund families with better recent performance tweet more actively, as well as families with lower volatility of returns.

The results in this section suggest that economies of scale are a key determinant of social media usage by asset managers, a feature that is shared with traditional advertisement (Sutton, 2007).

4 TWITTER ACTIVITY AND FUND FLOWS

In this section, we investigate whether mutual fund investors respond to the Twitter activity of asset management companies. More specifically, we study how the flows of money into the funds of a family are related to the family’s number of tweets and tone of those tweets in the previous month, controlling for fund performance and other well-documented flow determinants. To perform the analysis, we restrict the sample to funds of fund families that tweet at least once between January 2009 and December 2020.

Following the literature, we compute net flows to fund i between month t and month $t + \tau$ as the growth rate in total net assets net of the fund’s return:

$$\text{Flows}_{i,t+\tau} = \frac{\text{TNA}_{i,t+\tau} - \text{TNA}_{it}(1 + r_{i,t+\tau})}{\text{TNA}_{it}}, \quad (5)$$

where TNA_{it} is the total net assets of fund i at the end of month t , and $r_{i,t+\tau}$ is the fund’s compound monthly return between t and $t + \tau$. To minimize the impact of outliers - mostly small funds with large percentage of inflows or outflows - we follow the literature and winsorize flows at the 1% level.

Like [Sirri and Tufano \(1998\)](#), we allow for a non-linear flow-performance relationship. To model dependence on performance, we employ two different approaches. First, we define the variable Rank_{it} as the ranking of fund i ’s CAPM alpha in the 12-month period ending in month t against all other funds in the same Lipper category, normalized to be between $1/N$ (lowest performing fund) and 1 (highest performing fund), where N denotes the number of funds in the corresponding category and month. Second, we use objective-adjusted abnormal return (OAR) as in [Ha and Ko \(2019\)](#). We compute OAR_{it} by standardizing the 12-month CAPM alpha to have zero mean and unit standard deviation across all funds in the same investment category.

For both $\text{Performance}_{it} \in \{\text{Rank}_{it}, \text{OAR}_{it}\}$, we compute the following variables:

$$\begin{aligned} \text{Low Performance}_{it} &= \min(\text{Performance}_{it}, p20) \\ \text{Mid Performance}_{it} &= \min(\text{Performance}_{it} - \text{Low Performance}_{it}, p80 - p20) \\ \text{High Performance}_{it} &= \text{Performance}_{it} - \text{Mid Performance}_{it} - \text{Low Performance}_{it}, \end{aligned} \quad (6)$$

where $p20$, $p80$ denote the 20th and 80th percentiles, respectively, of either the cross-sectional distribution of performance rank or OAR.

To analyze the link between a fund family’s Twitter activity and subsequent flows, we estimate the

regression equation:

$$\begin{aligned}
\text{Flows}_{i,t+1} = & \gamma_0 + \gamma_1 \times \text{Twitter Activity}_{it} \\
& + \gamma_2 \times \text{Low Performance}_{it} + \gamma_3 \times \text{Mid Performance}_{it} + \gamma_4 \times \text{High Performance}_{it} \\
& + \Gamma \times X_{it} + \delta_t + \lambda_i + \mu_{cat} + \theta_{fam} + \nu_{i,t+1},
\end{aligned} \tag{7}$$

where Twitter Activity corresponds to either the Number of Tweets written by the fund family in a month in thousands, or the natural logarithm of one plus the number of tweets in thousands. $\text{Flows}_{i,t+1}$ is in %. Low, Mid, and High Performance are calculated using alternatively Rank or OAR based on the fund’s 12-month CAPM alpha, as in equation 6.¹¹ Following the large literature on the determinants of fund flows, the vector of lagged controls, X_{it} , includes the natural logarithm of the fund’s total net assets, the fund’s expense ratio, the fund’s age (log of months since inception), contemporaneous flows to funds in the same investment category excluding the instantaneous fund, return volatility in the previous 12 months, flows to the fund in the previous month. We also control for family size (log of assets under management) and family age (age of the family’s oldest fund). δ_t , λ_i , μ_{cat} , and θ_{fam} denote month, fund, investment category, and family fixed effects, respectively.¹² Finally, $\nu_{i,t+1}$ denotes the error term. We estimate equation (7) using pooled OLS and compute standard errors clustered at the month, fund family, and month-fund family levels.

Table 3 presents the results. In columns (1)-(3) we report results for OAR, and in columns (4)-(6) we use Ranks. Columns (1) and (4) are the baseline specifications without Twitter activity. In columns (2) and (5), we include Number of Tweets in thousands as the measure of Twitter activity. The coefficient on this variable is positive and statistically significant at the 1% level. To account for the possible non-linearity of the relation between flows and twitter activity, we also measure Twitter activity as the natural logarithm of 1 plus the number of tweets in thousands in columns (3) and (6).

Using objective-adjusted alpha as a measure of performance and the Number of Funds (columns 2 and 3), the estimated coefficients (0.50 and 0.69, respectively) imply that a one standard deviation increase in the number of tweets in a given month (69.59, as reported in Table 1) is associated with an increase in flows in the following month of approximately 0.035% ($= 0.50\% \times 0.06959$) when using the number of tweets to measure social media activity and 0.047% when using the variable in logs (evaluated at the average number of tweets per family and month, 29.5, as reported in Table 1).¹³ The results are al-

¹¹As explained above, we choose the CAPM to evaluate fund performance because this is the model that best explains fund flows (Barber et al., 2016; Berk and van Binsbergen, 2016).

¹²In our sample, some funds change investment categories and fund families through time.

¹³For the variable in logs, we approximate the change in flows as $\Delta(\text{Flows}) = 0.69\% \times \frac{d \ln(1 + \text{Number of Tweets})}{d \text{Number of Tweets}} \times \Delta(\text{Number of Tweets})$. Evaluating the derivative at the average number of tweets implies an increase in flows of $0.69\% \times \frac{1}{1+0.0295} \times 0.06959 = 0.047\%$ for a one standard deviation increase in tweets.

most identical if we use the estimated coefficients from the specification that models the flow-performance relation with performance ranks, rather than objective-adjusted returns (columns 5 and 6).

Another way to evaluate the size of the effect is to note that for an estimated coefficient of 0.50 on Number of Tweets (in thousands) and an average number of 29.5 tweets per family and month, and assuming a homogeneous effect of social media on flows across funds, Twitter activity generates net inflows of 0.01475% ($= 0.50 \times 0.0295$) each month per family. Since the average family in the Twitter sample manages assets worth USD 21,877.56 million, those inflows amount to USD 3.22 million each month for the average family.

Consistently with the literature, we find a convex relation between flows and performance for both ways of modelling non-linearity. Also, fund size, flows to funds in the same category, and volatility are all negatively associated with flows. Flows are persistent as evidenced by the positive and significant coefficient on lagged flows. Finally, younger funds capture more flows in percentage.

One concern in our analysis is that fund families could be tweeting about specific investment styles that experience higher flows as investors' preferences or expectations change through time. To tackle this potential concern, we repeat Table (3) removing time and investment category fixed effects, but including time-investment category fixed effect in the specification. Naturally, we remove the flows to the same category control from the specification. Results are presented in Table (4). Our main conclusions are unaltered, the estimated coefficients on the Number of Tweets and the log of the number of tweets become only slightly smaller relative to those in Table (3) and the standard errors are identical.

In the Internet Appendix we address a potential sample-selection bias in our tests due to the fact that we only include in the flow regressions those fund families that tweet at least once in our sample. In particular, we perform a two-step Heckman procedure where in the first step, we estimate a probit model of a fund belonging in the sample with fund and fund family characteristics, and then include the Mill's ratio as an explanatory variable in the second step. We find that the coefficient of positiveness remains unchanged, which alleviates the concern of a potential sample-selection bias in our results.

5 TONE OF TWEETS AND FUND FLOWS

In this section, we investigate whether the tone of the fund family's tweets and not just the number of tweets influence investor decisions. In Table (5) we repeat the regression analysis for the specification of equation (7) but including separately the number of positive tweets and the number of non-positive tweets (neutral + negative) in a given month, both in thousands. In columns (2) and (4) we find that the coefficient on the number of positive tweets in the month is positive and statistically significant while the

coefficient on the number of non-positive tweets is negative and statistically significant. The coefficient on the number of positive tweets is twice the size the coefficient on the number of tweets in Table (3).

A one standard deviation increase in the number of positive tweets is associated with a flow of USD 6.9 million to the average family ($1.09\% \times (29.04/1000) \times 21,799$). On the other hand, a one standard deviation increase in the number of non-positive tweets is associated with a decrease in flows of USD 9.098 million ($-1.7\% \times (24.55/1000) \times 21,799$).

One possible concern with these results is that tweets with different tones could be capturing different types of information. If differences in tone explain flows because they capture differences across topics, then we would observe no link between tone and flows if we restrict our attention to tweets of only one topic at a time. To perform this test, we count the number of positive, neutral, and negative tweets posted by each family and month in each one of the three main topics: Market Commentary; Financial Advice; and Customer Service. We then regress flows on the number of positive and non-positive (neutral and negative) tweets for each topic separately. The results of this test are reported in Table (6). We find that a larger number of positive (non-positive) tweets is associated with higher (lower) flows for each one of the three topics. However, both coefficients are statistically significant only for tweets classified as Market Commentary. For tweets classified as Customer Service, only the coefficient on positive flows is significant (at the 10% significance level). In the case of Financial Advice tweets, we find no statistically significant relation between the tone of the family's tweets and next-month flows. We conclude that the relation between positive tweets and subsequent flows survives after isolating tweets in different topics, although it remains statistically significant only for Market Commentary tweets.

These results also suggest that asset management tweets do not influence investors simply by providing a valuable customer service to their investors through social media. While we cannot rule out that possibility, the fact that Market Commentary tweets by themselves drive investor flows indicate that social media is more than just a channel for customer service for asset management firms.

Our results suggest that it is profitable for fund families to tweet massively in order to attract money from investors. However, the fact that not all fund families have a Twitter account nor tweet incessantly, particularly posts with a positive tone. This observation suggests that there must be some implicit costs associated with tweeting. We conjecture that reputational costs play a key role in determining the optimal twitter activity. For example, investors might evaluate ex-post the consistency between the tone of the tweets and respond more favorably when the sentiment of past tweets better reflects the current performance of the fund. They may also react negatively to tweets if the subsequent performance of the family's funds is mediocre.

To test this conjecture, We define variables of Consistency and Deception based on the overall tone

of the fund family’s communications at time $t - 1$ and the OAR of the fund at time t . More specifically, we define the following variables:

$$\begin{aligned}
\text{Consistency}(+) &= (\text{Positive Tweets}_{it} > \text{Negative Tweets}_{it}) \text{ and } (\text{OAR}_{it} > 0) \\
\text{Consistency}(-) &= (\text{Positive Tweets}_{it} < \text{Negative Tweets}_{it}) \text{ and } (\text{OAR}_{it} < 0) \\
\text{Consistency} &= \text{Consistency}(+) + \text{Consistency}(-) \\
\text{Deception} &= (\text{Positive Tweets}_{it} > \text{Negative Tweets}_{it}) \text{ and } (\text{OAR}_{it} < 0)
\end{aligned} \tag{8}$$

We then augment the regression equation (7) in Table (7) by adding these four variables in addition to Number of Tweets. Importantly, we continue to control for fund performance and other characteristics. In column (1) we include Consistency. In column (2) we include both Consistency(+) and Consistency(-). In column (3), we include only Deception. The estimated coefficient on Number of Tweets remains significant at the 10% level in columns (1) and (2) and at the 5% level in column (3). The estimated coefficients on our consistency measures suggest that investors prefer funds that have a performance consistent with the tone of the recent posts by their fund families. Funds with Consistency = 1 receive on average 0.2% higher net flows in a month than funds of the same performance and in families with similar Number of Tweets. However, in column (2) we observe that this relation is fully explained by the variable Consistency(+), whose estimated coefficient is identical to that on Consistency, whereas the coefficient on Consistency(-) is not statistically significant. Investors prefer funds in families with net positive communications if those funds subsequently outperform. Finally, the negative and significant coefficient on Deception in column (3) suggests that investors penalize funds that exhibit poor performance following net positive communications by their fund family. These funds receive on average 0.26 % fewer flows in a month than funds with similar performance and from families with similar Number of Tweets. Overall, these results are suggestive that indeed, there may be a cost to asset management companies posting positive tweets in terms of lower flows if funds’ subsequent performance is relatively poor.

6 ALTERNATIVE EXPLANATIONS

Our results so far suggest that more activity on Twitter by asset management firms is positively associated with subsequent flows to their funds, controlling for previously documented fund flow determinants as well as time-invariant fund and fund family characteristics. However, it is unclear whether social media communications themselves attract flows of money or, instead, they simply happen to be correlated with other asset management companies’ activities or events that determine flows. In the remainder of this

section, we consider and test three alternative explanations for a positive association between Twitter activity and mutual fund flows. We also perform a high-frequency analysis that allows us to isolate the impact of tweets on investor decisions from any confounders.

Our first alternative explanation is that asset management companies’ tweets are simply correlated with their marketing efforts. If posts on social media take place simultaneously with online, print, television advertising or with incentives to brokers, and we fail to control for such marketing activities, then social media posts are just a proxy for management companies’ marketing efforts, and not necessarily drivers of flows. To test this hypothesis, we perform two different tests. First, we control for the “virality” of the fund family’ tweets, i.e., the total number of “likes”, replies, and retweets (RT) for those tweets (divided by 10,000). The idea of the test is that if marketing—and not tweets—attract flows, then there is no reason why users’ response to tweets should be associated with flows. If, on the other hand, Twitter is an *independent channel* for flows, we would expect users’ reactions to be associated with subsequent flows. The estimation results are reported in Table 8. In column (1), we find that the estimated coefficient on the number of interactions with the fund family’s tweets is positive and statistically significant at the 1% level. The estimated coefficients on Number of Tweets (in 1,000) and virality (in 10,000), equal 0.49 and 0.03, respectively. In our sample, the average number of likes, replies and retweets per family and month equals 437, so users’ interactions with fund families’ tweets account for only a small fraction of the influence of social media on flows. To equate its influence, each tweet of the family would have to receive 163 likes, replies, and retweets ($= \frac{0.49}{0.03} \times \frac{10,000}{1,000}$).

In a second test, we control explicitly for the fund family’s marketing efforts, as proxied by funds’ expenses used for advertising, marketing, and compensation to brokers. More specifically, we first compute for each share class the sum of 12b-1 fees and annualized loads. To compute annualized loads we divide the maximum loads of the share class by 3.¹⁴ We then compute the asset-weighted average of marketing expenses across all share classes (within our sample) in the family, as a proxy for the marketing efforts of the fund’s family. To allow for a non-linear effect of marketing on flows, we compute terciles of this variable for every month at the family level and include two indicator variables for the second and third terciles. The results are shown in columns (2) and (3). In column (2) we include only our proxies for marketing expenses. In column (3), we include also the interaction between the number of tweets and marketing expenses. We find that marketing expenses are positively associated with fund flows, consistently with research on mutual fund marketing (Jain and Wu, 2000; Cronqvist, 2006; Gallaher et al., 2015). This effect is economically significant. Funds whose fund family has marketing expenses in the top (second) tercile of the distribution receive in average 0.27% (0.11%) higher flows compared to funds in the bottom

¹⁴The average holding period of equity mutual funds is estimated to be between 2 and 3 years (O’Neal, 2004).

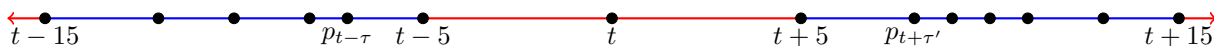
tercile. However, the coefficient on Number of Tweets remains positive and significant. In column (3) we find that the interaction between Number of Tweets and the top tercile dummy is positive and statistically significant. The estimated coefficients suggest that funds from fund families in the top tercile of marketing expenses have a marginal benefit of tweeting of 0.63 compared to 0.38 for the second tercile and -0.37 in the bottom tercile.

These findings give further credence to the notion that tweets are not just proxying for marketing efforts through other channels. Our results are consistent with asset managers' social media strategy being a separate channel from marketing, although both channels for attracting flows seem to reinforce each other, at least for families that invest a lot in marketing.

Third, we consider the possibility that tweets convey information that is already publicly available and relevant for future fund flows. Accordingly, Twitter activity itself may have no impact on fund flows and the estimated association is the consequence of omitted variable bias. To explore this possibility, we reestimate the flow regression equation controlling for the following information: i) a change in the fund's management company in the previous month; ii) a change in the fund's portfolio manager or manager team in the previous month; iii) an indicator variable if the fund is in the Top quintile of its OAR distribution. Changes in the fund's management could be perceived by investors as a positive signal of future returns (Khorana, 2001). Stellar performance may attract disproportionate flows as they capture attention from the media (Kaniel and Parham, 2017). Results in Table 9 indicate that the association between Number of Tweets and fund flows remains positive, similar in magnitude, and statistically significant at the 1% level after controlling for changes in management, or stellar performance.

Of course, the set of events we have considered in Table 9 is not exhaustive. Without a source of exogenous variation in Twitter activity, isolating the effect of tweets on open-end mutual fund flows from confounders is not possible. Instead, we propose to identify the effect of asset management companies tweets on investor decisions by exploiting the high-frequency nature of social media communications in combination with intraday ETF price data. The analysis is based on the idea that asset management companies' positive tweets increase the demand not just for their open-end mutual funds but also for their ETFs. Since ETF shares are traded in an exchange, we expect a family's positive tweets to put upward pressure on the price of its ETFs. The main advantage of using ETFs over open-end mutual funds for the purpose of this analysis is that ETFs are traded intraday, which allows us to identify the effect of tweeting activity on prices using the high-frequency methodology of Bianchi et al. (2023) and Bianchi et al. (2024). More specifically, we study whether an ETF's price changes in a very short time window around the time when the ETF's family tweets. As explained by those authors, this high-frequency approach allows for clean identification of the effect of tweets on prices with the identifying assumption being that

over such a short window of time no other relevant information affecting the security’s price is released. To perform the test, we follow closely [Bianchi et al. \(2024\)](#). First, we use the timestamp in Eastern Time of every tweet in our sample to determine two 10-minute windows: one from 15 to 5 minutes *before* the tweet and another one from 5 to 15 minutes *after* the tweet. Second, for each one of the family’s ETFs, we identify the price of the last trade in the [-15,-5] minute window and the price of the first trade in the [+5,+15] minute window as depicted below:¹⁵



Finally, we compute the change in log prices and regress it on the tweet’s Vader Score.¹⁶ Following [Bianchi et al. \(2023\)](#) and [Bianchi et al. \(2024\)](#) we include fund fixed effects. That is, we estimate the regression equation:

$$\Delta p_{i,t} = a + b \times \text{Vader Sentiment}_{i,t} + \delta_i + \epsilon_{i,t},$$

where $\Delta p_{i,t}$ is the change in ETF i ’s log price between the last trade of the [-15,-5] minute window and the first trade of the [+5,+15] minute window around a tweet by i ’s family taking place at time t .

Estimation results are reported in [Table 10](#). In column (1) we study the effect of family tweets on its ETF share prices between 5 minutes before the tweet and 5 minutes after each tweet. The estimated coefficient is positive and statistically significant at the 5% significance level, which indicates that family tweets with a positive tone push up the prices of their ETFs relative to neutral tweets. Based on the estimated coefficient, a tweet with a VADER sentiment of one increases ETF share prices by 0.09 basis points. This result is strong evidence in favor of the hypothesis that tweets themselves drive flows of money to the fund family’s products, and do not simply proxy for omitted actions of fund families or other events.

Although this result is consistent with tweets increasing the demand for the fund family’s products, it could also be that tweets correctly predict positive information about the ETF’s underlying assets. For instance, asset managers tweet good news about the market that the ETF tracks. If that is true, we would expect the effect on prices to be temporary, as any price deviations from fundamentals (the ETF’s net asset value) will be arbitrated away by authorized participants. In this case, the ETF shares would quickly revert to the net asset value, although the ETF would grow in size. If, on the other hand, tweets anticipate an increase in the value of the ETF’s assets, then the increase in price would be permanent.

To distinguish between both possibilities, we consider 10-minute windows that start later after the tweet.

¹⁵We exclude the observation from the analysis if either the last price in the window before the tweet or the first price in the window after the tweet have zero associated volume in our dataset

¹⁶We winsorize price changes at the 1% level.

More specifically, we consider eight different 10-minute windows starting at $t + 15$, and finishing at $t + 85$ and recalculate $\Delta p_{i,t}$ accordingly always using the same *before* price. In column (2) we move forward the window after the tweet by 10 minutes and estimate an even larger and statistically significant marginal impact of 0.2 basis points. The larger coefficient could be explained by the fact that we are giving investors more time to react. Also, the impact of one tweet is likely amplified over time through retweets and cross-posting in social media. When we study price changes 35 minutes after the tweet (column 4), the estimated coefficient has a value of 0.65 basis points. After this time the marginal coefficient decreases, and in column (8) it has a value of 0.08 and it is no longer significant, so the effect of a tweet is completely reversed 75 minutes after the tweet. This finding is consistent with tweets increasing the demand for the fund family’s products and not with tweets anticipating increases in the value of the portfolio’s assets.

Since the individual impact of a tweet on ETF prices might not be symmetric across positive and negative tweets we repeat the analysis by computing the positive and negative component of the VADER Score separately. Each variable is defined on the interval $[0, 1]$ and they are not simultaneously positive. The results are presented in Table (11), we observe that the marginal impact of a positive tweet disappears after 55 minutes, while negative tweets have a positive impact in the first 15 minutes and their impact reverses after 45 minutes and disappears after 75 minutes.

One possible concern about the high-frequency tests is that they may not be sufficient to isolate the effect of tweets on ETF prices. For instance, on a day when favorable news is released, ETF prices might increase throughout the day as investors digest the information. If asset management companies tweet about that event, it could appear that ETF prices respond favorably to those tweets, whereas, in reality, the prices are incorporating the news with some lag. To alleviate this concern, we exploit the high-frequency nature of our data and include day fixed effects in our specification. The day fixed effects absorb the impact on ETF prices of any events or news that affect prices on a given day. The results of this high-frequency analysis are reported in the Internet Appendix. The estimated coefficients are positive and statistically significant for the VADER sentiment score of the tweets. When regressing ETF price changes on the positive and negative scores separately, we continue to find that ETF prices increase with positive tweets and decrease with negative tweets.

In the Internet Appendix, we also include fund-month fixed effects to control for all observable and unobservable ETF characteristics that change from month to month and which may influence ETF price changes. The results corroborate that ETF prices respond to the tone of asset management companies’ tweets.

Taken together, the results in this section provide strong support to the hypothesis that the social media communications of asset management companies are a distinct channel for influencing investor deci-

sions that operates independently from marketing. In other words, asset managers’ social media strategy does not simply mirror their marketing strategy. Social media is not a mere vehicle for customer service, either. Finally, social media in asset management does not simply echo publicly available information. Instead, our findings suggest that social media is a powerful new tool in the hands of asset managers with a larger potential to influence investor decisions.

7 ECONOMIC MECHANISMS

In this section, we investigate the mechanism through which the social media activity of asset management companies attracts investors’ money to the funds they manage. More specifically, we consider four possible hypotheses. First, asset management companies use social media to reduce search costs for investors. Second, asset management companies’ tweets increase the attention that investors pay to their managed funds. Third, asset management companies use social media to convey to investors performance-relevant information that is *not publicly available*. Fourth, fund families use social media as a tool to persuade investors, i.e., to create uninformative messages that exploit investor biases.

We start by testing whether asset management companies use social media to reduce search costs for investors. This can be achieved by facilitating investors’ access to information that is already available but difficult to locate for investors, such as information about fund offerings, fees, past performance, or investment strategy. An example of such tweets is shown in Figure (8) where *BlackRock* explains the strategy of its Global Allocation Fund. Building on the work of [Sirri and Tufano \(1998\)](#) and [Hortaçsu and Syverson \(2004\)](#), we posit that a reduction in search costs increases flows of money from new investors, as more investors are willing to pay the cost of searching for the family’s funds. But lower search costs also decrease the flow-performance sensitivity, particularly in the high-performance region, as potential investors facing lower search costs are willing to consider more of the family’s funds and not just those at the very top of the performance distribution ([Huang et al., 2007](#)).

To test these theoretical predictions, we investigate how Inflows and Outflows and their relation to performance changes with Number of Tweets. More specifically, we define Inflows and Outflows for fund i in month $t + 1$ as:

$$\begin{aligned} \text{Inflows}_{i,t+1} &= \frac{\text{New Sales}_{i,t+1}}{\text{TNA}_{i,t}} \\ \text{Outflows}_{i,t+1} &= \frac{\text{Redeemed Cash}_{i,t+1}}{\text{TNA}_{i,t}} \end{aligned}$$

As argued by [Ha and Ko \(2019\)](#) inflows and outflows are simultaneously determined by investors’ rebal-

ancing strategies. To tackle the mutual dependence between inflows and outflows we follow closely the methodology proposed by [Ha and Ko \(2019\)](#). In particular, we first estimate the following OLS regressions for inflows and outflows separately:

$$\begin{aligned} \text{Inflows}_{i,t+1} &= a + \sum_{s=0}^{11} b_s \times \text{Inflows}_{i,t-s} + c \times X_{it} + \nu_{i,t+1}, \\ \text{Outflows}_{i,t+1} &= a + \sum_{s=0}^{11} b_s \times \text{Outflows}_{i,t-s} + c \times X_{it} + \nu_{i,t+1}, \end{aligned}$$

where X_{it} contains the same controls used in the flow regressions. We then regress inflows and outflows on the fitted values on outflows and inflows, respectively, and keep the residuals. Finally, residual inflows and residual outflows are regressed on Number of Tweets and performance (Rank and OAR).

In [Table 12](#), we show estimation results for inflows in columns (1) to (4) and for outflows in columns (5) to (8). In the odd columns, we include no interactions of Number of Tweets with performance. In the even columns, we include both Number of Tweets and their interactions with the measures of performance that define the piecewise-linear function. The coefficient of Inflows on Number of Tweets is positive and significant for both OAR and Rank (Columns 1 and 3). With respect to the interactions, we find that the interaction between Number of Tweets and performance in the high-performance region is positive significant for OAR but it is not significant for Rank. Therefore, the results are consistent with the prediction that tweets attract money from new investors, but not with the prediction that the relation between flows of money from new investors and performance becomes flatter in the high-performance region. Although these results are not supportive of the search-cost hypothesis, they should be taken with caution. Tweeting is an endogenous decision. In particular, fund families could be strategically posting in times when their overall performance is high. Such good general performance would by itself make the family’s funds more visible to outside investors, so a larger number of tweets by the family would have little impact on the relation between performance and new money, particularly in the high-performance range.

The results of [Table 12](#) are also informative about the attention hypothesis. Models of rational inattention predict that increases in investor attention to stocks result in more retail investors considering both purchases and sales of stocks, although the increase in purchases by retail investors should be higher than the increase in sales due to costs and constraints affecting short sales ([Barber and Odean, 2008](#); [Da et al., 2011](#)). The increase in inflows following a higher number of tweets that we observe in [Table 12](#) is consistent with this story. However, we also find that a higher number of tweets *reduces* outflows across funds, which contradicts the theoretical prediction. Also, when investors are paying attention, underperforming funds in families with good overall performance should experience more withdrawals, as investors

would discriminate more between the worst funds in the family and others. In Table 12 (columns 6 and 8), we find exactly the opposite result: Outflows become *less* sensitive to performance in the low performance region (for both measures of performance) following an increase in the number of tweets by the family. More specifically, while outflows decrease with higher performance, a higher number of tweets make the slope of the flow-performance relation less negative in the low-performance region: A one standard deviation increase in the number of tweets is associated with an *increase* in the slope of the outflow-performance relation of about 13% the size of the slope for families that do not tweet. In sum, the findings from the analysis of inflows and outflows are not supportive of the attention hypothesis, either.

We now move on to testing the hypothesis that asset management companies use social media to convey to investors performance-relevant information that is not publicly available. The model of Dumitrescu and Gil-Bazo (2016) of strategic communication by asset managers predicts that asset management companies will choose to communicate only privately observed information that is favorable with respect to future fund performance. Since in equilibrium, such communications are truthful, not previously known, and favorable to the fund, they should have a positive impact on flows. But such communications should also possess predictive ability with respect to future fund performance beyond and above public information. To test this prediction of the theory, we regress one-month ahead performance on Number of Tweets while controlling for past performance and fund and family characteristics that have been documented in the literature to predict performance. One difficulty that arises with this test is the fact that net performance is partially determined by investors' reaction to Twitter Activity. In particular, if there are diseconomies of scale in asset management, fund performance will deteriorate as money flows into the fund (Berk and Green, 2004). To account for diseconomies of scale, we control also for fund size and past flows. More specifically, we estimate:

$$\hat{\alpha}_{i,t+1} = \rho_0 + \rho_1 \times \alpha_{i,t-3:t} + \rho_2 \times \text{Number of Tweets (1,000)}_{i,t} + P \times X_{it} + \delta_t + \lambda_i + \mu_{cat} + \theta_{fam} + \nu_{i,t+1}, \quad (9)$$

where $\alpha_{i,t-3:t}$ is the fund's abnormal return in the previous three months. X_{it} is a vector of control variables that includes: fund size; expense ratio; flows (in month t); portfolio turnover; 12-month return volatility; an indicator variable that equals one if the fund charges loads; fund age; family size and family age. We also control for our Retail indicator since this variable may correlate both with Twitter activity and performance. δ_{t+1} , λ_i , μ_{cat} , and θ_{fam} denote time, fund, investment category, and family fixed effects, respectively. Standard errors are clustered at the family, month, and family-month levels.

Table 13 shows the estimation results. In column (1), we use CAPM alpha both as the dependent

variable and as a control (in this case, measured over the previous three months). In column (2), our measure of performance is the [Fama and French \(1993\)](#) 3-factor alpha. In column (3), we use Carhart’s (1997) four-factor alpha. In column (4), we use [Fama and French \(2015\)](#) five-factor alpha. And in column (5) we estimate alpha using the five-factor model of [Fama and French \(2015\)](#) augmented with momentum. The coefficients on Number of Tweets are insignificant in all specifications. Therefore, regardless of the factor model we use to measure performance, Number of Tweets of fund i in month t does not predict whether the fund will outperform in month $t + 1$. These findings do not support the hypothesis that positive tweets reduce asymmetric information about future fund performance.

It could be argued that not all categories of tweets are equally informative about future performance. We exploit our topic classification and repeat the analysis separately for tweets in each topic. That is, we compute the variable Number of Tweets counting tweets within one topic classification at a time. We then regress flows on these three variables and the rest of potential flow determinants. The estimation results, are included in the Internet Appendix. We find no evidence that tweets of any topic predict future fund performance for any of the measures of performance considered.

Although these results do not support the hypothesis that tweets may contain valuable information about future performance, it should be noted that not all categories of tweets are equally informative. To test whether at least some tweets predict performance, we exploit our topic classification and repeat the analysis separately for tweets in each topic. That is, we compute the variable Number of Tweets counting tweets within one topic classification at a time. We then regress flows on these three variables and the rest of potential flow determinants. According to the estimation results, which we include in the Internet Appendix, tweets in all three topics show no predictive power with respect to future fund performance.

It could also be that the average predictability across all funds in the family masks a stronger link between Twitter activity and subsequent performance for specific groups of funds. To explore this possibility, we identify tweets that make reference to a particular investment category. Given that our sample consists of US diversified equity mutual funds, we focus on the investment styles in the 3x3 Morningstar matrix (size by value). Using a dictionary of words commonly used to refer to specific investment styles, we find very few such tweets. For instance, the most commonly referenced investment style in tweets according to our dictionary is “small caps”. On average, a family posts only 0.09 small-cap tweets in a given month, as opposed to almost 30 total tweets. Only 0.24% of all the tweets in our sample contain words that make reference to small caps. For other styles, that fraction ranges between 0.11% and 0.02% of all tweets. Nevertheless, we count the number of tweets in each style and perform the regression analysis restricting the sample to mutual funds in that style (as classified by Lipper). The estimation results are reported in the Appendix. We find no evidence that the number of tweets in a given investment predicts

performance of funds in that category. However, the test probably has little power given the sparsity of the variable of interest.

We then study whether Twitter activity predicts future performance for funds that benefit the most from tweets.¹⁷ One way of testing this conjecture is to first determine the fund characteristics that are associated with higher flows due to social media activity and then test whether the number of tweets predicts higher performance for funds with those characteristics. In Table 14, we run flow regressions that include interactions between the number of tweets and several fund characteristics. We find that funds charging lower expense ratios, funds in larger and older families, and funds with higher recent past flows benefit the most from family tweets. Given these results, in Table 15, we regress CAPM alpha on the number of tweets in the previous month and interactions with those four characteristics. We find that only the interaction with expense ratio is statistically significant: A higher number of tweets predicts significantly higher alpha for cheaper funds. In the Internet Appendix, we repeat the analysis for other measures of performance.

Overall, we find no evidence that social media activity predicts superior fund performance, with the exception of low-fee funds.

Finally, we test the persuasion hypothesis, i.e., the notion that asset managers influence investor decisions through uninformative messages. To investigate the persuasion hypothesis directly, we build on the work of Mullainathan and Shleifer (2005) and Mullainathan et al. (2008) and test their theoretical prediction that persuasion requires that the message conveyed to investors be adapted to different circumstances. More specifically, in times of low investor sentiment, asset managers will want to associate their products to safety and trust, whereas in times of high investor sentiment, asset managers will favor the idea of seizing investment opportunities. Consequently, successful persuasion will lead investors towards safer products such as money market or bond funds in low-sentiment markets and towards riskier products, such as equity funds, in high-sentiment markets. Since our sample consists of only equity funds, if social media activity is used to persuade investors, we would expect Number of Tweets to be more strongly associated with flows in times of high investor sentiment.

However, when testing for this hypothesis, we need to be aware of a potential confounder. Market sentiment may correlate with a higher fraction of purchases coming from retail investors and that retail investors are generally more sensitive to social media activity. Therefore, we need to control for both our Retail dummy and its interaction with Number of Tweets.

¹⁷We have also identified individual funds that have recently benefitted the most (top tercile of flows in their family) when their asset management companies have tweeted a lot (top tercile of the Number of Tweets distribution). We have found no evidence that a higher Number of Tweets in the previous month predicts higher performance for those funds. However, the performance of those funds, regardless of the Number of Tweets, tends to decline.

We first test whether flows from retail investors are more responsive to asset management companies' tweets. To do this, we regress flows on Number of Tweets, the Retail dummy, the interaction between these two variables, and controls. The results, reported in Table 16 confirm that net purchases of mutual funds by retail investors are indeed more sensitive to Number of Tweets. In fact, once we include the interaction between Retail and Number of Tweets, the main effect of Number of Tweets is no longer significant, which suggests that the influence of social media on mutual fund flows is driven by retail investors.

Having confirmed the higher sensitivity of retail investor flows to tweets, we test whether the link between Number of Tweets and Flows is stronger in times of high market sentiment, controlling for the retail dummy and its interaction with Number of Tweets. The results are shown in Table 17. First, we regress flows on only the Sentiment measure of Baker and Wurgler (2006).¹⁸ The coefficient on Sentiment is not significant. In column (2) we include Number of Tweets and its interaction with Sentiment and find that only the coefficient on the interaction is positive and statistically significant (at the 10% level). In column (3), we repeat the analysis replacing Number of Tweets with the number of Positive Tweets, which we would expect to be the most persuasive in times of high investor sentiment. In this case, both the coefficient on Positive Tweets and its interaction with Sentiment are positive and significant. These results confirm that Twitter activity is more effective in times of high market sentiment.

If we add the Retail dummy and its interaction with Number of Tweets (column 4), we find that the coefficient on the interaction between Number of Tweets and Sentiment does not change substantially and retains its significance. When we repeat the analysis replacing the number of tweets with the number of positive tweets, we reach similar conclusions. Therefore, the increased link between social media activity and flows during times of high investor sentiment does not seem to be driven by more retail investor activity during those periods.

Taken together, the results in this section do not support that asset management firms use social media to reduce search costs for their funds. They do not support, either, the hypothesis that Twitter activity by fund families draws investors' attention towards their funds. We also find no evidence that asset management companies use social media to convey private performance-relevant information to investors. However, the results seem consistent with asset managers using social media to persuade investors by creating uninformative messages that attract flows and slow withdrawals.

¹⁸We download from Jeffrey Wurgler's website a measure of investor sentiment based on the first principal component of five sentiment proxies as well as a different measure, where each of the proxies has first been orthogonalized with respect to a set of six macroeconomic indicators.

8 CONCLUSIONS

We unveil Twitter as a novel channel for attracting flows to mutual funds. Asset management companies' social media communications predict subsequent flows of money to their equity funds. This association between social media activity and future flows is not explained by social media activity being correlated with fund families' marketing efforts. Instead, social media strategies are distinct from marketing strategies although both strategies reinforce each other. Social media in asset management is not merely a vehicle for customer service or for reiterating public announcements, either. Our high-frequency analysis exploiting ETF trade data gives credence to the idea that tweets themselves impact investor decisions.

An analysis of inflows and outflows does not support a search cost or an attention mechanism driving our results. The fact that tweets have not predictive ability with respect to future performance also suggests that asset managers do not use social media to alleviate information asymmetries.

Instead, our results support the persuasion hypothesis. Tweets, particularly those with a more positive tone, attract more money from investors to equity funds. Tweets are most effective for retail investors and when investor sentiment is higher, consistently with asset managers creating uninformative managers to reinforce investors' perceptions of increased opportunities in capital markets.

REFERENCES

- Abis, S. and Lines, A. (2022). Do mutual funds keep their promises? *Available at SSRN 3851861*.
- Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4):1645–1680.
- Bandiera, O., Prat, A., Hansen, S., and Sadun, R. (2020). CEO Behavior and Firm Performance. *Journal of Political Economy*, 128(4):1325–1369.
- 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.
- Barber, B. M. and Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, 21(2):785–818.
- Berk, J. B. and Green, R. (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112(6):1269–1295.
- Berk, J. B. and van Binsbergen, J. (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics*, 119(1):1–23.
- Bertomeu, J. and Marinovic, I. (2016). A theory of hard and soft information. *The Accounting Review*, 91(1):1–20.
- Bianchi, F., Gómez-Cram, R., Kind, T., and Kung, H. (2023). Threats to central bank independence: High-frequency identification with twitter. *Journal of Monetary Economics*, 135:37–54.
- Bianchi, F., Gómez-Cram, R., and Kung, H. (2024). Using Social Media to Identify the Effects of Congressional Viewpoints on Asset Prices. *The Review of Financial Studies*, 37(7):2244–2272.
- Blankespoor, E., Miller, G. S., and White, H. D. (2014). The role of dissemination in market liquidity: Evidence from firms’ use of Twitter™. *The Accounting Review*, 89(1):79–112.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.
- Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.
- Bubb, R. and Catan, E. M. (2022). The party structure of mutual funds. *The Review of Financial Studies*, 35(6):2839–2878.
- Cade, N. L. (2018). Corporate social media: How two-way disclosure channels influence investors. *Accounting, Organizations and Society*, 68-69:63–79.
- Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82.
- Christoffersen, S. E. K., Evans, R., and Musto, D. K. (2013). What Do Consumers’ Fund Flows Maximize? Evidence from Their Brokers’ Incentives. *Journal of Finance*, 68(1):201–235.
- Cookson, J. A., Fox, C., Gil-Bazo, J., Imbet, J. F., and Schiller, C. (2023). Social media as a bank run catalyst. *Available at SSRN 4422754*.
- Cookson, J. A., Lu, R., Mullins, W., and Niessner, M. (2024a). The social signal. *Journal of Financial Economics*, 158:103870.
- Cookson, J. A., Mullins, W., and Niessner, M. (2024b). Social media and finance. *Available at SSRN 4806692*.

- Cookson, J. A. and Niessner, M. (2023). Investor disagreement: Daily measures from social media. *Available at SSRN 4529594*.
- Cornelli, F., Kominek, Z., and Ljungqvist, A. (2013). Monitoring managers: Does it matter? *Journal of Finance*, 68(2):431–481.
- Cronqvist, H. (2006). Advertising and portfolio choice. *Working Paper*.
- Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *The journal of finance*, 66(5):1461–1499.
- Dumitrescu, A. and Gil-Bazo, J. (2016). Information and investment under uncertainty. *Economics Letters*, 148(C):17–22.
- Dumitrescu, A., Gil-Bazo, J., and Zhou, F. (2022). Defining greenwashing. *Working Paper*.
- Dye, R. A. and Sridhar, S. (2004). Reliability-relevance trade-offs and the efficiency of aggregation. *Journal of Accounting Research*, 42(1):51–88.
- Evans, R. B. and Fahlenbrach, R. (2012). Institutional investors and mutual fund governance: Evidence from retail–institutional fund twins. *The Review of Financial Studies*, 25(12):3530–3571.
- Evans, R. B. and Sun, Y. (2020). Models or stars: The role of asset pricing models and heuristics in investor risk adjustment. *The Review of Financial Studies*.
- Fama, E. and French, K. (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.
- Gallaher, S. T., Kaniel, R., and Starks, L. T. (2015). Advertising and mutual funds: From families to individual funds. *Working Paper*.
- Gholampour, V. and van Wincoop, E. (2019). Exchange rate disconnect and private information: What can we learn from euro-dollar tweets? *Journal of International Economics*, 119:111–132.
- Gil-Bazo, J. and Ruiz-Verdú, P. (2009). The relation between price and performance in the mutual fund industry. *Journal of Finance*, 64(5):2153–2183.
- Gu, C. and Kurov, A. (2020). Informational role of social media: Evidence from twitter sentiment. *Journal of Banking and Finance*, 121:105969.
- Ha, Y. and Ko, K. (2019). Misspecifications in the fund flow-performance relationship. *Journal of Financial Intermediation*, 38:69–81.
- Hansen, S., McMahon, M., and Prat, A. (2018). Transparency and deliberation within the fomc: A computational linguistics approach*. *The Quarterly Journal of Economics*, 133(2):801–870.
- Hillert, A., Niessen-Ruenzi, A., and Ruenzi, S. (2023). Mutual fund shareholder letters: Flows, performance, and managerial behavior. *Working Paper*.
- Hortaçsu, A. and Syverson, C. (2004). Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. *Quarterly Journal of Economics*, 119(2):403–456.
- Huang, J., Wei, K. D., and Yan, H. (2007). Participation costs and the sensitivity of fund flows to past performance. *Journal of Finance*, 62(3):1273–1311.
- Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225.

- Jain, P. C. and Wu, J. S. (2000). Truth in mutual fund advertising: Evidence on future performance and fund flows. *Journal of Finance*, 55(2):937–958.
- Kaniel, R. and Parham, R. (2017). Wsj category kings—the impact of media attention on consumer and mutual fund investment decisions. *Journal of Financial Economics*, 123(2):337–356.
- Khorana, A. (2001). Performance changes following top management turnover: Evidence from open-end mutual funds. *Journal of Financial and Quantitative Analysis*, 36(3):371–393.
- Kim, O. and Verrecchia, R. (1991). Trading volume and price reactions to public announcements. *Journal of Accounting Research*, 29(2):302–321.
- Leippold, M. and Rueegg, R. (2020). How rational and competitive is the market for mutual funds? *Review of Finance*, 24(3):579–613.
- Long, S., Lucey, B., Xie, Y., and Yarovaya, L. (2023). “i just like the stock”: The role of reddit sentiment in the gamestop share rally. *Financial Review*, 58(1):19–37.
- Loughran, T. and McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4):1187–1230.
- Mullainathan, S., Schwartzstein, J., and Shleifer, A. (2008). Coarse Thinking and Persuasion. *The Quarterly Journal of Economics*, 123(2):577–619.
- Mullainathan, S. and Shleifer, A. (2005). Persuasion in finance. *Working Paper*.
- O’Neal, E. S. (2004). Purchase and redemption patterns of us equity mutual funds. *Financial Management*, pages 63–90.
- Pastor, L., Stambaugh, R., and Taylor, L. A. (2015). Scale and skill in active management. *Journal of Financial Economics*, 116(1):23–45.
- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., and Mozetič, I. (2015). The effects of twitter sentiment on stock price returns. *PloS one*, 10(9):e0138441.
- Roussanov, N., Ruan, H., and Wei, Y. (2021). Marketing mutual funds. *The Review of Financial Studies*, 34(6):3045–3094.
- Sirri, E. R. and Tufano, P. (1998). Costly search and mutual fund flows. *Journal of Finance*, 53(5):1589–1622.
- Sutton, J. (2007). *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration*, volume 1 of *MIT Press Books*. The MIT Press.
- You, W., Guo, Y., and Peng, C. (2017). Twitter’s daily happiness sentiment and the predictability of stock returns. *Finance Research Letters*, 23:58–64.

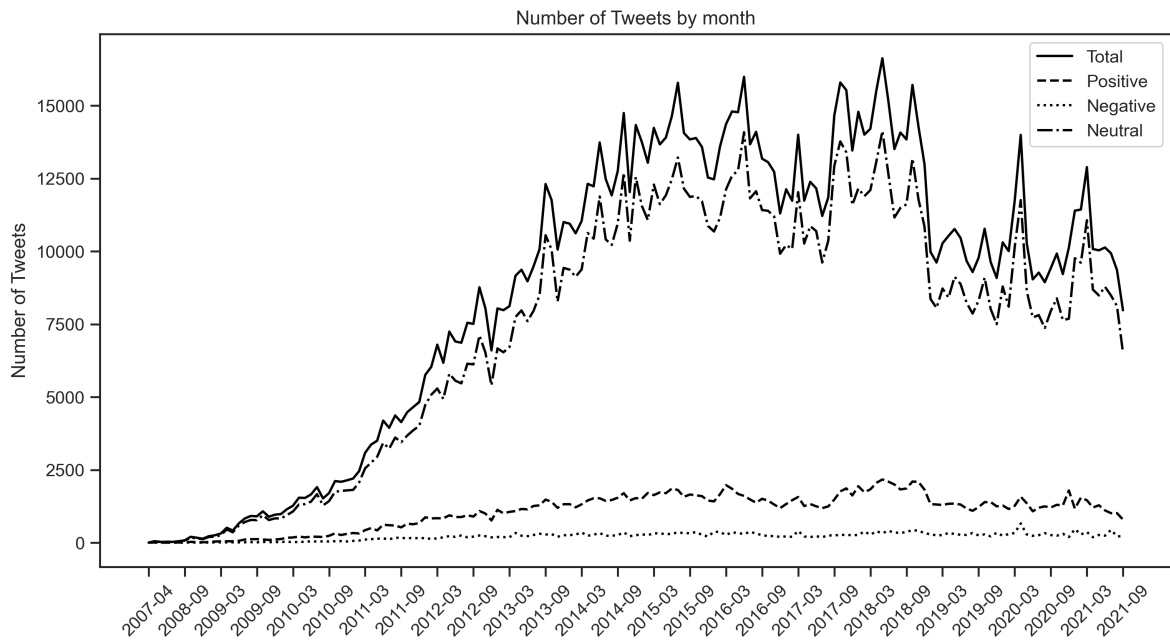


Figure 1: Evolution of the number tweets by all fund families through time. The figure shows the number of tweets posted by all fund families per month. The solid black line shows the total number of tweets. The dashed, dotted, and dot-dashed lines represent the number of tweets classified as positive, negative, and neutral respectively.



Northern Trust
@NorthernTrust



Robust growth is a challenge in #India, #Brazil and #Indonesia, says our Chief Economist: bit.ly/Gzywvf

9:33 PM · Oct 1, 2013

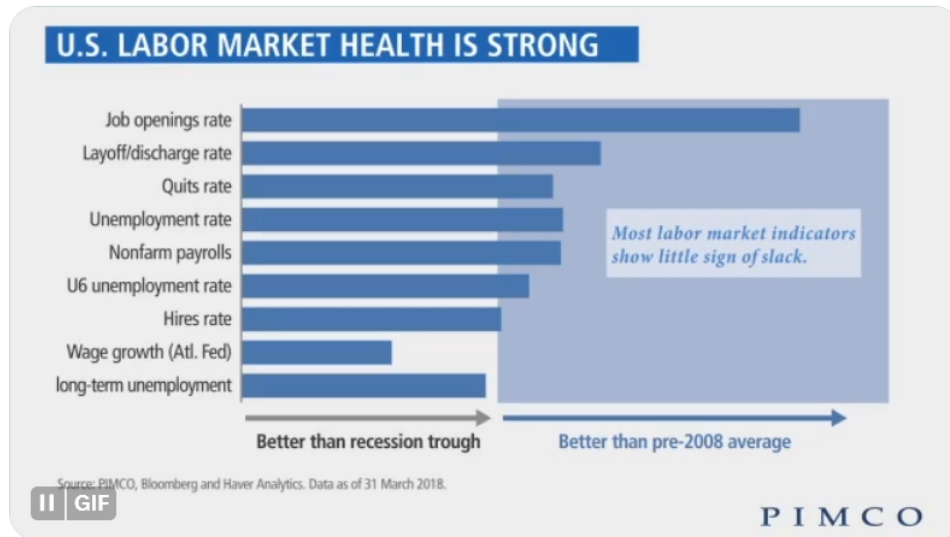
Figure 5: Example of a tweet classified as negative with a score of 0.994. The tweet was written by asset management company Northern Trust (@NorthernTrust) on October 1 2013.



PIMCO
@PIMCO



The U.S. labor market is one reason for PIMCO's solid growth outlook for the year ahead pim.co/vzehk



4:05 PM · May 26, 2018

Figure 6: Example of a tweet classified as positive with a score of 1. The tweet was written by asset management company PIMCO (@PIMCO) on May 26, 2018.

361 **361 Capital, a division of Hamilton Lane**
@361Capital

\$SPX +0.4% to new ATH

German IFO 92.6 bte to Feb high

July New Home Sales 901k bte,
and supply falls to 4.0 months.

Richmond Fed Mfg +18 bte

Aug Cons Conf 84.8 wte

Bonds fall: ↑ econ data + ↑ supply

Best: Oil, Ag, EM, **\$QQQ**, Ind Metals

Worst: Energy, Utes, Bonds, US\$

10:41 PM · Aug 25, 2020

Figure 7: Example of a tweet with topic loadings of 0.99 for market commentary, 0.0 for customer service and 0.0 for financial advice.



One of the cheapest times to [#travel](#) is weeks away. Visit [#MyRetirement](#) for tips on travel deals in [#retirement](#) bit.ly/1yAgFNu

11:01 PM · Nov 20, 2014

Figure 8: Example of a tweet with topic loadings 0.01 for market commentary, 0.98 for customer service and 0.0 for financial advice.



A5: Know if your HDHP is HSA-qualified otherwise you will have a high deductible without the benefit of putting tax-free money away to help pay for current and future medical expenses. [#HSAsHelp](#)
[#HSAsHelpChat](#)

7:21 PM · Oct 15, 2019

Figure 9: Example of a tweet with topic loadings 0.13 for market commentary, 0.07 for customer service and 0.79 for financial advice.

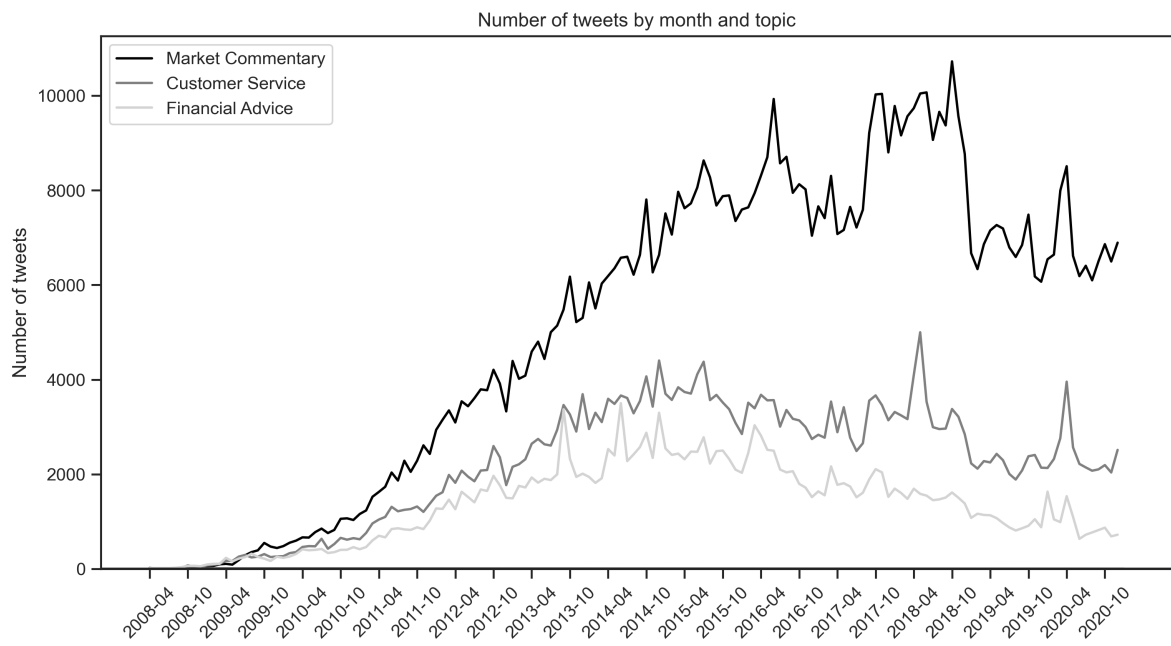


Figure 10: Evolution of Tweets by Topic.

Table 1: Descriptive Statistics, Fund and Fund Family characteristics

This table contains summary statistics of fund and fund family characteristics for two samples. The first sample, Twitter Subsample, consists of fund families managing U.S. equity funds that have tweeted at least once from January 2009 to December 2020. The Full Sample includes all families managing U.S. equity funds in the same period. The first set of rows show descriptive statistics for variables computed at the fund-month level, while the second set of rows show descriptive statistics for variables computed at the fund family level. Average front (back) loads are the average of loads charged by the fund across all investment (redemption) intervals. The tenure is computed as the number of months that the manager has been consecutively in office in months. Flows to the same category in percentage is computed as the flows to all funds in the same Lipper investment classification excluding the instantaneous fund. All remaining variables are defined in Section (2).

| Variable | Twitter Subsample | | | | | | Full Sample | | | | | |
|---------------------------------|-------------------|---------|----------|--------|----------|--------|-------------|--------|----------|--------|----------|--------|
| | mean | median | s.d. | p1 | p99 | N | mean | median | s.d. | p1 | p99 | N |
| Fund Level | | | | | | | | | | | | |
| Positive Tweets | 22.07 | 2.22 | 51.73 | 0.00 | 256.48 | 530921 | | | | | | |
| Negative Tweets | 1.97 | 0.13 | 3.58 | 0.00 | 16.28 | 530921 | | | | | | |
| Neutral Tweets | 13.41 | 2.00 | 21.74 | 0.00 | 94.00 | 530921 | | | | | | |
| Total Tweets | 59.88 | 24.00 | 93.92 | 0.00 | 405.00 | 530921 | | | | | | |
| Flows (%) | -0.39 | -0.60 | 5.66 | -18.33 | 18.65 | 525906 | -0.39 | -0.60 | 5.66 | -18.33 | 18.65 | 525906 |
| Age (months) | 161.18 | 139.00 | 120.35 | 31.00 | 709.00 | 530921 | 161.18 | 139.00 | 120.35 | 31.00 | 709.00 | 530921 |
| Expense Ratio | 0.01 | 0.01 | 0.01 | 0.00 | 0.02 | 530921 | 0.01 | 0.01 | 0.01 | 0.00 | 0.02 | 530921 |
| Total Net Assets (USD millions) | 892.29 | 149.40 | 3383.27 | 15.70 | 13358.40 | 530921 | 892.29 | 149.40 | 3383.27 | 15.70 | 13358.40 | 530921 |
| Average Front Load | 0.04 | 0.04 | 0.00 | 0.02 | 0.05 | 104472 | 0.04 | 0.04 | 0.00 | 0.02 | 0.05 | 104472 |
| Average Back Load | 0.01 | 0.01 | 0.01 | 0.00 | 0.02 | 57547 | 0.01 | 0.01 | 0.01 | 0.00 | 0.02 | 57547 |
| Turnover | 0.71 | 0.47 | 1.99 | 0.00 | 4.81 | 530921 | 0.71 | 0.47 | 1.99 | 0.00 | 4.81 | 530921 |
| Tenure (months) | 104.82 | 94.00 | 62.62 | 10.00 | 290.00 | 465801 | 104.82 | 94.00 | 62.62 | 10.00 | 290.00 | 465801 |
| Flow to Same Category (%) | -0.14 | -0.32 | 1.19 | -2.61 | 4.22 | 525897 | -0.14 | -0.32 | 1.19 | -2.61 | 4.22 | 525897 |
| Fraction of Positive Tweets | 0.49 | 0.46 | 0.22 | 0.00 | 1.00 | 296900 | 0.49 | 0.46 | 0.22 | 0.00 | 1.00 | 296900 |
| CAPM alpha (%) | -0.15 | -0.15 | 2.12 | -5.82 | 5.59 | 530921 | -0.15 | -0.15 | 2.12 | -5.82 | 5.59 | 530921 |
| Fund Family Level | | | | | | | | | | | | |
| Positive Tweets | 9.08 | 0.00 | 29.04 | 0.00 | 131.50 | 22839 | | | | | | |
| Negative Tweets | 1.26 | 0.00 | 3.36 | 0.00 | 16.90 | 22839 | | | | | | |
| Neutral Tweets | 8.38 | 0.00 | 19.73 | 0.00 | 93.00 | 22839 | | | | | | |
| Total Tweets | 29.50 | 1.00 | 69.59 | 0.00 | 318.00 | 22839 | | | | | | |
| Flows (%) | -0.01 | -0.01 | 0.07 | -0.13 | 0.16 | 22757 | -0.00 | -0.01 | 0.15 | -0.14 | 0.18 | 65277 |
| Age (months) | 348.30 | 246.00 | 293.52 | 32.00 | 1075.00 | 22839 | 264.22 | 203.00 | 231.69 | 31.00 | 1035.00 | 65677 |
| Volatility (%) | 4.47 | 4.06 | 2.23 | 1.05 | 11.72 | 22839 | 4.50 | 4.08 | 2.25 | 0.92 | 11.47 | 65677 |
| Funds with Loads (%) | 92.99 | 100.00 | 13.03 | 41.18 | 100.00 | 22839 | 94.50 | 100.00 | 12.93 | 48.28 | 100.00 | 65677 |
| Expense Ratio | 0.01 | 0.01 | 0.00 | 0.00 | 0.02 | 22839 | 0.01 | 0.01 | 0.00 | 0.00 | 0.02 | 65677 |
| Avg. CAPM alpha 12 months (%) | -0.00 | -0.00 | 0.01 | -0.02 | 0.02 | 22839 | -0.00 | -0.00 | 0.01 | -0.02 | 0.02 | 65677 |
| Number of Funds | 25.69 | 6.00 | 45.17 | 1.00 | 199.00 | 22839 | 16.03 | 2.00 | 77.25 | 1.00 | 150.00 | 65677 |
| Number of Investment Categories | 6.45 | 4.00 | 6.74 | 1.00 | 28.00 | 22839 | 3.98 | 2.00 | 5.21 | 1.00 | 25.00 | 65677 |
| Total Net Assets (USD millions) | 21799.40 | 1578.10 | 82152.02 | 17.60 | 4.8e+05 | 22839 | 11296.54 | 390.20 | 62090.32 | 16.50 | 2.1e+05 | 65677 |
| Unique fund families | | | | | | 269 | | | | | | 848 |

Table 2: Determinants of Twitter Activity

This table shows estimation results for regressions of Twitter activity on family characteristics. Columns (1) to (3) provide results of running a regression of a dummy variable that equals 1 if a fund family tweets at least once in a month on lagged fund family characteristics. Finally columns (4) to (6) regress the logarithm of one plus the number of tweets posted by a fund family in a given month on fund family characteristics lagged one month only for fund families that tweet at least once in our sample. Standard errors clustered at the month level are presented in columns (1) to (6). Finally, ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level, respectively.

| | Number of Tweets > 0 | | | log(1+Number of Tweets) | | |
|--------------------------------|----------------------|----------------------|----------------------|-------------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| log(TNA (USD millions)) | 0.023*** (0.002) | | | 0.120*** (0.013) | | |
| log(Number of Funds) | | 0.034*** (0.004) | | | -0.025 (0.019) | |
| log(Number of Inv. Categories) | | | 0.043*** (0.004) | | | -0.067*** (0.024) |
| log(1+Age (months)) | -0.063*** (0.009) | -0.062*** (0.009) | -0.061*** (0.009) | -0.152*** (0.052) | -0.074 (0.049) | -0.065 (0.049) |
| Avg. CAPM alpha 12 months (%) | 0.643*** (0.222) | 0.757*** (0.219) | 0.757*** (0.219) | 2.972*** (1.081) | 3.501*** (1.135) | 3.436*** (1.130) |
| Expense Ratio | -0.186 (0.494) | -0.870* (0.480) | -1.185** (0.492) | 5.176 (4.676) | -2.873 (4.599) | -3.122 (4.554) |
| Volatility (%) | -0.010*** (0.001) | -0.010*** (0.001) | -0.010*** (0.001) | -0.063*** (0.010) | -0.064*** (0.010) | -0.065*** (0.010) |
| Observations | 64706 | 64706 | 64706 | 22591 | 22591 | 22591 |
| Adjusted R^2 (%) | 68.65 | 68.62 | 68.62 | 70 | 69.91 | 69.91 |
| Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Fund Family FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Sample | Full | Full | Full | Twitter | Twitter | Twitter |

Table 3: Twitter Activity and Flows

This table shows estimation results for regressions of mutual fund flows in percentage points on the number of tweets, the logarithm of one plus the number of tweets, and control variables. Standard errors clustered at the month, fund family, and month-fund family level, are shown in parentheses. All variables are defined in Section (2). Finally ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level, respectively.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---------------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Number of Tweets (1,000) | | 0.50*** (0.16) | | | 0.49*** (0.16) | |
| log(1+Number of Tweets (1,000)) | | | 0.69*** (0.21) | | | 0.67*** (0.21) |
| Low OAR | 0.33*** (0.03) | 0.33*** (0.03) | 0.33*** (0.03) | | | |
| Mid OAR | 0.67*** (0.02) | 0.67*** (0.02) | 0.67*** (0.02) | | | |
| High OAR | 0.79*** (0.04) | 0.79*** (0.04) | 0.79*** (0.04) | | | |
| Low Rank | | | | 2.73*** (0.23) | 2.72*** (0.23) | 2.72*** (0.23) |
| Mid Rank | | | | 1.67*** (0.06) | 1.67*** (0.06) | 1.67*** (0.06) |
| High Rank | | | | 8.42*** (0.43) | 8.42*** (0.43) | 8.42*** (0.43) |
| log(TNA (USD millions)) | -0.64*** (0.03) | -0.64*** (0.03) | -0.64*** (0.03) | -0.65*** (0.03) | -0.65*** (0.03) | -0.65*** (0.03) |
| Expense Ratio | 5.08** (2.36) | 5.06** (2.34) | 5.03** (2.34) | 5.27** (2.46) | 5.25** (2.45) | 5.22** (2.44) |
| Volatility (%) | -9.88*** (1.14) | -9.87*** (1.14) | -9.87*** (1.14) | -8.00*** (1.13) | -7.99*** (1.13) | -7.99*** (1.13) |
| Family Age (log) | -0.20** (0.09) | -0.20** (0.09) | -0.20** (0.09) | -0.21** (0.09) | -0.21** (0.09) | -0.21** (0.09) |
| Fund Age (log) | -1.90*** (0.10) | -1.89*** (0.10) | -1.89*** (0.10) | -1.94*** (0.10) | -1.94*** (0.10) | -1.94*** (0.10) |
| Family Size (log) | 0.27*** (0.05) | 0.27*** (0.05) | 0.27*** (0.05) | 0.28*** (0.05) | 0.28*** (0.05) | 0.28*** (0.05) |
| Past Flows | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) | 0.15*** (0.01) | 0.15*** (0.01) | 0.15*** (0.01) |
| Flow to Same Category (%) | 0.44*** (0.01) | 0.44*** (0.01) | 0.44*** (0.01) | 0.43*** (0.01) | 0.43*** (0.01) | 0.43*** (0.01) |
| Constant | 10.87*** (0.90) | 10.82*** (0.90) | 10.81*** (0.90) | 10.31*** (0.90) | 10.26*** (0.90) | 10.25*** (0.90) |
| Observations | 453572 | 453572 | 453572 | 453599 | 453599 | 453599 |
| Adjusted R^2 (%) | 13.21 | 13.22 | 13.22 | 13.16 | 13.17 | 13.17 |
| Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Inv. Category FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Fund FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Fund Family FE | Yes | Yes | Yes | Yes | Yes | Yes |

Table 4: Twitter Activity and Flows (alternative specification)

This table shows estimation results for regressions of mutual fund flows in percentage points on the number of tweets, the logarithm of one plus the number of tweets, and control variables. Standard errors clustered at the month, fund family, and month-fund family level, are shown in parentheses. All variables are defined in Section (2). Finally ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level, respectively.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|-----------------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Number of Tweets (1,000) | | 0.45*** (0.16) | | | 0.45*** (0.16) | |
| log(1+Number of Tweets (1,000)) | | | 0.63*** (0.21) | | | 0.63*** (0.21) |
| Low OAR | 0.32*** (0.03) | 0.32*** (0.03) | 0.32*** (0.03) | | | |
| Mid OAR | 0.65*** (0.02) | 0.65*** (0.02) | 0.65*** (0.02) | | | |
| High OAR | 0.81*** (0.04) | 0.81*** (0.04) | 0.81*** (0.04) | | | |
| Low Rank | | | | 2.69*** (0.23) | 2.69*** (0.23) | 2.69*** (0.23) |
| Mid Rank | | | | 1.66*** (0.06) | 1.66*** (0.06) | 1.66*** (0.06) |
| High Rank | | | | 8.74*** (0.43) | 8.74*** (0.43) | 8.74*** (0.43) |
| log(TNA (USD millions)) | -0.67*** (0.03) | -0.67*** (0.03) | -0.67*** (0.03) | -0.68*** (0.03) | -0.68*** (0.03) | -0.68*** (0.03) |
| Expense Ratio | 4.16* (2.23) | 4.11* (2.22) | 4.08* (2.21) | 4.59* (2.36) | 4.54* (2.34) | 4.51* (2.34) |
| Volatility (%) | -21.84*** (1.83) | -21.76*** (1.83) | -21.76*** (1.82) | -17.48*** (2.04) | -17.40*** (2.03) | -17.40*** (2.03) |
| Family Age (log) | -0.18** (0.09) | -0.18** (0.09) | -0.18** (0.09) | -0.19** (0.09) | -0.19** (0.09) | -0.19** (0.09) |
| Fund Age (log) | -1.84*** (0.11) | -1.84*** (0.11) | -1.84*** (0.11) | -1.86*** (0.11) | -1.86*** (0.11) | -1.86*** (0.11) |
| Family Size (log) | 0.23*** (0.05) | 0.23*** (0.05) | 0.23*** (0.05) | 0.24*** (0.05) | 0.24*** (0.05) | 0.24*** (0.05) |
| Past Flows | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) |
| Constant | 11.57*** (0.92) | 11.52*** (0.92) | 11.51*** (0.92) | 10.76*** (0.92) | 10.71*** (0.92) | 10.70*** (0.92) |
| Observations | 453499 | 453499 | 453499 | 453499 | 453499 | 453499 |
| Adjusted R^2 (%) | 14.09 | 14.09 | 14.09 | 14.05 | 14.05 | 14.05 |
| Time FE | No | No | No | No | No | No |
| Inv. Category FE | No | No | No | No | No | No |
| Fund FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Fund Family FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Time FE \times Inv. Category FE | Yes | Yes | Yes | Yes | Yes | Yes |

Table 5: Twitter Tone and Flows

This table shows estimation results for regressions of mutual fund flows in percentage points on the number of positive and non positive (neutral + negative) tweets, and control variables. Lagged Flows are computed equivalently in the last month, last three, last six, and last twelve months. Standard errors clustered at the month, fund family, and month-fund family level, are shown in parentheses. All variables are defined in Section (2). Finally ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level, respectively.

| | (1) | (2) | (3) | (4) |
|-----------------------------|--------------------|--------------------|--------------------|--------------------|
| Positive Tweets (1,000) | | 1.09*** (0.35) | | 1.01*** (0.34) |
| Non-Positive Tweets (1,000) | | -1.70** (0.72) | | -1.52** (0.72) |
| Low OAR | 0.33*** (0.03) | 0.33*** (0.03) | | |
| Mid OAR | 0.67*** (0.02) | 0.66*** (0.02) | | |
| High OAR | 0.79*** (0.04) | 0.79*** (0.04) | | |
| Low Rank | | | 2.73*** (0.23) | 2.73*** (0.23) |
| Mid Rank | | | 1.67*** (0.06) | 1.66*** (0.06) |
| High Rank | | | 8.42*** (0.43) | 8.40*** (0.43) |
| log(TNA (USD millions)) | -0.64*** (0.03) | -0.64*** (0.03) | -0.65*** (0.03) | -0.65*** (0.03) |
| Expense Ratio | 5.08** (2.36) | 5.22** (2.37) | 5.27** (2.46) | 5.39** (2.47) |
| Volatility (%) | -9.88*** (1.14) | -9.80*** (1.14) | -8.00*** (1.13) | -7.93*** (1.13) |
| Family Age (log) | -0.20** (0.09) | -0.20** (0.09) | -0.21** (0.09) | -0.21** (0.09) |
| Fund Age (log) | -1.90*** (0.10) | -1.89*** (0.10) | -1.94*** (0.10) | -1.93*** (0.10) |
| Family Size (log) | 0.27*** (0.05) | 0.27*** (0.05) | 0.28*** (0.05) | 0.28*** (0.05) |
| Past Flows | 0.14*** (0.01) | 0.14*** (0.01) | 0.15*** (0.01) | 0.15*** (0.01) |
| Flow to Same Category (%) | 0.44*** (0.01) | 0.44*** (0.01) | 0.43*** (0.01) | 0.43*** (0.01) |
| Constant | 10.87*** (0.90) | 10.86*** (0.90) | 10.31*** (0.90) | 10.30*** (0.90) |
| Observations | 453572 | 453572 | 453599 | 453599 |
| Adjusted R^2 (%) | 13.21 | 13.22 | 13.16 | 13.17 |
| Time FE | Yes | Yes | Yes | Yes |
| Inv. Category FE | Yes | Yes | Yes | Yes |
| Fund FE | Yes | Yes | Yes | Yes |
| Fund Family FE | Yes | Yes | Yes | Yes |

Table 6: Twitter Tone, Topics, and Flows

This table shows estimation results for regressions of mutual fund flows in percentage points on the number of positive and non-positive (neutral+negative) tweets that are classified as market commentary, customer service or financial advice. Control variables are defined in Section (2). Standard errors, clustered at the month-family level, are shown in parentheses. Finally ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level, respectively.

| | (1) | (2) | (3) |
|--|--------------------|--------------------|--------------------|
| Positive Tweets (Market Commentary) (1,000) | 2.52*** (0.68) | | |
| Non-Positive Tweets (Market Commentary) 1000 | -6.98*** (2.00) | | |
| Positive Tweets (Customer Service) (1,000) | | 3.19* (1.89) | |
| Non-Positive Tweets (Customer Service) 1000 | | -8.88 (7.08) | |
| Positive Tweets (Financial Advice) (1,000) | | | 9.08 (9.71) |
| Non-Positive Tweets (Financial Advice) 1000 | | | -11.29 (9.24) |
| Low OAR | 0.40*** (0.09) | 0.40*** (0.09) | 0.40*** (0.09) |
| Mid OAR | 0.66*** (0.08) | 0.66*** (0.08) | 0.66*** (0.08) |
| High OAR | 0.78*** (0.10) | 0.78*** (0.10) | 0.79*** (0.10) |
| log(TNA (USD millions)) | -2.19*** (0.25) | -2.19*** (0.25) | -2.19*** (0.25) |
| Expense Ratio | 12.65 (10.23) | 12.59 (10.24) | 12.38 (10.21) |
| Volatility (%) | -8.80* (5.01) | -8.96* (5.01) | -8.97* (5.01) |
| Family Age (log) | -0.05 (0.23) | -0.08 (0.23) | -0.08 (0.23) |
| Fund Age (log) | -1.04*** (0.29) | -1.06*** (0.29) | -1.07*** (0.29) |
| Family Size (log) | 0.85*** (0.23) | 0.86*** (0.23) | 0.86*** (0.23) |
| Past Flows | 0.13*** (0.01) | 0.13*** (0.01) | 0.13*** (0.01) |
| Flow to Same Category (%) | 0.53*** (0.04) | 0.53*** (0.04) | 0.52*** (0.04) |
| Constant | 8.21*** (2.86) | 8.27*** (2.87) | 8.29*** (2.85) |
| Observations | 452509 | 452509 | 452509 |
| Adjusted R^2 (%) | 1.75 | 1.74 | 1.75 |
| Time FE | Yes | Yes | Yes |
| Inv. Category FE | Yes | Yes | Yes |
| Fund FE | Yes | Yes | Yes |
| Fund Family FE | Yes | Yes | Yes |

Table 7: Twitter Tone, Consistency of Information, and Flows

This table shows estimation results for regressions of mutual fund flows for one, three, six, and twelve months in percentage points on the number of tweets, the consistency of Twitter communications, and control variables. Consistency(+) is a dummy variable that takes the value of one if at month $t - 1$ the fund family tweets more positive than negative tweets, and at month t the OAR of the fund is positive. Consistency(-) is a dummy variable that takes the value of one if at month $t - 1$ the fund family tweets more negative than positive tweets, and at month t the OAR of the fund is negative. Consistency is defined as Consistency(+) + Consistency(-). Deception is a dummy variable that takes the value of one if at month $t - 1$ the fund family tweets more positive than negative tweets, and at month t the OAR of the fund is negative. Lagged Flows are computed equivalently in the last month, last three, last six, and last twelve months. Standard errors, clustered at the month, fund family, and month-fund family levels are shown in parentheses. All variables are defined in Section (2). ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level, respectively.

| | (1) | (2) | (3) |
|---------------------------|--------------------|--------------------|--------------------|
| Number of Tweets (1,000) | 0.67* (0.38) | 0.67* (0.38) | 0.78** (0.38) |
| Consistency | 0.21* (0.11) | | |
| Consistency(+) | | 0.21* (0.11) | |
| Consistency(-) | | 0.10 (0.18) | |
| Deception | | | -0.26** (0.12) |
| Low OAR | 0.41*** (0.09) | 0.41*** (0.09) | 0.42*** (0.09) |
| Mid OAR | 0.56*** (0.12) | 0.56*** (0.12) | 0.53*** (0.11) |
| High OAR | 0.80*** (0.10) | 0.80*** (0.10) | 0.80*** (0.10) |
| log(TNA (USD millions)) | -2.19*** (0.25) | -2.19*** (0.25) | -2.20*** (0.25) |
| Expense Ratio | 12.30 (10.17) | 12.29 (10.17) | 12.31 (10.18) |
| Volatility (%) | -9.08* (5.01) | -9.09* (5.01) | -9.09* (5.02) |
| Family Age (log) | -0.09 (0.23) | -0.09 (0.23) | -0.06 (0.23) |
| Fund Age (log) | -1.05*** (0.29) | -1.05*** (0.29) | -1.05*** (0.29) |
| Family Size (log) | 0.87*** (0.23) | 0.87*** (0.23) | 0.85*** (0.23) |
| Past Flows | 0.13*** (0.01) | 0.13*** (0.01) | 0.13*** (0.01) |
| Flow to Same Category (%) | 0.53*** (0.04) | 0.53*** (0.04) | 0.53*** (0.04) |
| Constant | 8.21*** (2.87) | 8.22*** (2.87) | 8.34*** (2.87) |
| Observations | 452509 | 452509 | 452509 |
| Adjusted R^2 (%) | 1.75 | 1.75 | 1.75 |
| Time FE | Yes | Yes | Yes |
| Inv. Category FE | Yes | Yes | Yes |
| Fund FE | Yes | Yes | Yes |
| Fund Family FE | Yes | Yes | Yes |

Table 8: Tweets' Virality, Marketing, and Flows

This table shows estimation results for regressions of mutual fund flows in percentage points on the number of tweets, the total number of likes plus replies plus retweets across all tweets posted by a family, and two dummies that indicate if a fund family marketing expenses correspond to the last two terciles of the distribution in a month. Control variables are defined in Section (2). Standard errors clustered at the month, fund family, and month-fund family level, are shown in parentheses. All variables are defined in Section (2). Finally ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level, respectively.

| | (1) | (2) | (3) |
|--|--------------------|--------------------|--------------------|
| Number of Tweets (1,000) | 0.49*** (0.16) | 0.37** (0.16) | -0.37 (0.36) |
| (Likes + Replies + Retweets) (10,000) | 0.03*** (0.01) | | |
| Marketing Expenses Family (T=2) | | 0.11** (0.05) | 0.09* (0.05) |
| Number of Tweets (1,000) × Marketing Expenses Family (T=2) | | | 0.75 (0.50) |
| Marketing Expenses Family (T=3) | | 0.27*** (0.07) | 0.21*** (0.08) |
| Number of Tweets (1,000) × Marketing Expenses Family (T=3) | | | 1.00** (0.41) |
| Low OAR | 0.33*** (0.03) | 0.33*** (0.03) | 0.33*** (0.03) |
| Mid OAR | 0.66*** (0.02) | 0.67*** (0.02) | 0.67*** (0.02) |
| High OAR | 0.79*** (0.04) | 0.79*** (0.04) | 0.79*** (0.04) |
| log(TNA (USD millions)) | -0.64*** (0.03) | -0.64*** (0.03) | -0.64*** (0.03) |
| Expense Ratio | 5.06** (2.34) | 5.05** (2.34) | 5.14** (2.35) |
| Volatility (%) | -9.87*** (1.14) | -9.82*** (1.14) | -9.78*** (1.14) |
| Family Age (log) | -0.20** (0.09) | -0.21** (0.09) | -0.20** (0.09) |
| Fund Age (log) | -1.89*** (0.10) | -1.89*** (0.10) | -1.89*** (0.10) |
| Family Size (log) | 0.27*** (0.05) | 0.27*** (0.05) | 0.27*** (0.05) |
| Past Flows | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) |
| Flow to Same Category (%) | 0.44*** (0.01) | 0.44*** (0.01) | 0.44*** (0.01) |
| Constant | 10.82*** (0.90) | 10.77*** (0.89) | 10.76*** (0.89) |
| Observations | 453572 | 453572 | 453572 |
| Adjusted R^2 (%) | 13.22 | 13.22 | 13.22 |
| Time FE | Yes | Yes | Yes |
| Inv. Category FE | Yes | Yes | Yes |
| Fund FE | Yes | Yes | Yes |
| Fund Family FE | Yes | Yes | Yes |

Table 9: Twitter Activity, Fundamental Information, and Flows

This table shows estimation results for regressions of mutual fund flows in percentage points on the number of tweets, three variables that reflect changes in fundamentals and control variables. Change Family is a dummy variable that takes the value of one if there is a change in the variable mgmt_cd in CRSP between two consecutive months, Change Manager is a dummy variable that takes the value of one if there is a change in the variable mgr_name in CRSP between two consecutive months accounting for team managed funds and non-alphanumeric characters. Top Quintile OAR is a dummy variable that takes the value of one if the fund's OAR is in the top quintile of the OAR distribution. Control variables are defined in Section (2). Standard errors clustered at the month, fund family, and month-fund family level, are shown in parentheses. All variables are defined in Finally ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level, respectively.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Number of Tweets (1,000) | | 0.51*** (0.16) | | 0.50*** (0.16) | | 0.51*** (0.16) |
| Change Family | -0.15 (0.20) | -0.15 (0.20) | | | | |
| Change Manager | | | -0.21*** (0.07) | -0.21*** (0.07) | | |
| Top Quintile OAR | | | | | 0.13*** (0.04) | 0.13*** (0.04) |
| Low OAR | 0.33*** (0.03) | 0.33*** (0.03) | 0.33*** (0.03) | 0.33*** (0.03) | 0.34*** (0.03) | 0.34*** (0.03) |
| Mid OAR | 0.67*** (0.02) | 0.67*** (0.02) | 0.67*** (0.02) | 0.67*** (0.02) | 0.62*** (0.02) | 0.62*** (0.02) |
| High OAR | 0.79*** (0.04) | 0.79*** (0.04) | 0.79*** (0.04) | 0.79*** (0.04) | 0.72*** (0.05) | 0.72*** (0.05) |
| log(TNA (USD millions)) | -0.64*** (0.03) | -0.64*** (0.03) | -0.64*** (0.03) | -0.64*** (0.03) | -0.64*** (0.03) | -0.64*** (0.03) |
| Expense Ratio | 5.08** (2.36) | 5.06** (2.34) | 5.09** (2.36) | 5.06** (2.35) | 5.12** (2.37) | 5.09** (2.35) |
| Volatility (%) | -9.88*** (1.14) | -9.87*** (1.14) | -9.89*** (1.14) | -9.88*** (1.14) | -9.73*** (1.14) | -9.72*** (1.14) |
| Family Age (log) | -0.20** (0.09) | -0.20** (0.09) | -0.20** (0.09) | -0.20** (0.09) | -0.20** (0.09) | -0.20** (0.09) |
| Fund Age (log) | -1.90*** (0.10) | -1.89*** (0.10) | -1.90*** (0.10) | -1.89*** (0.10) | -1.90*** (0.10) | -1.90*** (0.10) |
| Family Size (log) | 0.27*** (0.05) | 0.27*** (0.05) | 0.27*** (0.05) | 0.27*** (0.05) | 0.27*** (0.05) | 0.27*** (0.05) |
| Past Flows | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) |
| Flow to Same Category (%) | 0.44*** (0.01) | 0.44*** (0.01) | 0.44*** (0.01) | 0.44*** (0.01) | 0.44*** (0.01) | 0.44*** (0.01) |
| Constant | 10.87*** (0.90) | 10.82*** (0.90) | 10.87*** (0.90) | 10.82*** (0.90) | 10.92*** (0.90) | 10.86*** (0.90) |
| Observations | 453572 | 453572 | 453572 | 453572 | 453572 | 453572 |
| Adjusted R^2 (%) | 13.21 | 13.22 | 13.22 | 13.22 | 13.22 | 13.22 |
| Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Inv. Category FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Fund FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Fund Family FE | Yes | Yes | Yes | Yes | Yes | Yes |

Table 10: High Frequency Analysis: Twitter Tone

This table presents OLS estimates for the impact of the sentiment of tweets posted by fund families in our sample on their ETFs' price changes. In all regressions the dependent variable Δp is the log change in the ETF's price between the last trade of the $[-15, -5]$ minute window and the first trade in the time window presented in parenthesis in every column. Price changes are winsorized at the 1% level and are expressed in basis points. Vader Sentiment is defined in the interval $[-1, 1]$. Standard errors clustered at the fund, month, and fund-month levels are presented in parenthesis. Finally ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level.

| | (1) $\Delta p(5, 15)$ | (2) $\Delta p(15, 25)$ | (3) $\Delta p(25, 35)$ | (4) $\Delta p(35, 45)$ | (5) $\Delta p(45, 55)$ | (6) $\Delta p(55, 65)$ | (7) $\Delta p(65, 75)$ | (8) $\Delta p(75, 85)$ |
|--------------------|--------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Vader Sentiment | 0.09** (0.04) | 0.20*** (0.05) | 0.40*** (0.06) | 0.65*** (0.08) | 0.51*** (0.09) | 0.27*** (0.09) | 0.19* (0.10) | 0.08 (0.11) |
| Constant | -0.17*** (0.02) | -0.23*** (0.02) | -0.20*** (0.03) | -0.16*** (0.04) | -0.12*** (0.04) | -0.25*** (0.05) | -0.28*** (0.05) | -0.20*** (0.06) |
| Observations | 1789392 | 1756007 | 1723952 | 1672237 | 1669693 | 1625233 | 1547799 | 1516605 |
| Adjusted R^2 (%) | .07 | .07 | .07 | .07 | .07 | .07 | .07 | .07 |
| Fund FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

Table 11: High Frequency Analysis: Positive and Negative Tweets

This table presents OLS estimates for the impact of the sentiment of tweets posted by fund families in our sample on their ETFs' price changes. In all regressions the dependent variable Δp is the log change in the ETF's price between the last trade of the $[-15, -5]$ minute window and the first trade in the time window presented in parenthesis in every column. Price changes are winsorized at the 1% level and are expressed in basis points. Vader Positive is defined as the Vader Sentiment if it is positive and zero otherwise. Vader Negative is defined as the negative of the Vader Sentiment if it is negative and zero otherwise. Standard errors clustered at the fund, month, and fund-month levels are presented in parenthesis. Finally ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level.

| | (1) $\Delta p(5, 15)$ | (2) $\Delta p(15, 25)$ | (3) $\Delta p(25, 35)$ | (4) $\Delta p(35, 45)$ | (5) $\Delta p(45, 55)$ | (6) $\Delta p(55, 65)$ | (7) $\Delta p(65, 75)$ | (8) $\Delta p(75, 85)$ |
|--------------------|--------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Vader Positive | 0.48*** (0.04) | 0.39*** (0.06) | 0.54*** (0.07) | 0.75*** (0.09) | 0.52*** (0.10) | 0.12 (0.11) | 0.08 (0.12) | 0.03 (0.13) |
| Vader Negative | 1.03*** (0.11) | 0.32** (0.15) | 0.01 (0.18) | -0.37* (0.21) | -0.49** (0.23) | -0.70*** (0.26) | -0.49* (0.28) | -0.25 (0.31) |
| Constant | -0.34*** (0.02) | -0.31*** (0.03) | -0.26*** (0.03) | -0.21*** (0.04) | -0.13*** (0.04) | -0.19*** (0.05) | -0.23*** (0.06) | -0.17*** (0.06) |
| Observations | 1789392 | 1756007 | 1723952 | 1672237 | 1669693 | 1625233 | 1547799 | 1516605 |
| Adjusted R^2 (%) | .08 | .07 | .07 | .07 | .07 | .07 | .07 | .07 |
| Fund FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

Table 12: Twitter Activity, Inflows and Outflows

This table shows estimation results for regressions of residual inflows and residual outflows (both in %) on the number of tweets and fund performance. In a first stage (not reported) Inflows (Out- flows) are regressed on 12 lags of the variable and controls: Size, Flows to the same category, Volatility, Expense Ratio, Age, one-month lagged flows to the fund, Family size, and Family age. In a second stage (not reported) we regress Inflows (Outflows) on the fitted values of Outflows (Inflows) estimated in the first stage. Fitted residuals from the regression are then regressed on the number of tweets and the performance variables. OLS robust standard errors are presented in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level, respectively.

| | Inflows | | | | Outflows | | | |
|--------------------------|-------------------|-------------------|--------------------|-------------------|--------------------|--------------------|---------------------|---------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Number of Tweets (1,000) | 2.51** (1.22) | -2.61 (2.63) | 2.35* (1.20) | 1.62 (2.10) | -1.49*** (0.56) | 5.46 (4.23) | -1.47*** (0.56) | -10.89*** (4.20) |
| × Low OAR | | -1.88 (1.97) | | | | 8.57* (4.57) | | |
| × Mid OAR | | 3.92 (2.68) | | | | -0.17 (1.04) | | |
| × High OAR | | 5.25** (2.27) | | | | 1.14 (1.12) | | |
| × Low Rank | | | | -19.67 (20.29) | | | | 50.17** (23.80) |
| × Mid Rank | | | | 15.46 (11.76) | | | | 0.01 (2.51) |
| × High Rank | | | | -16.99 (54.85) | | | | 11.55 (12.13) |
| Low OAR | 0.11 (0.20) | 0.15 (0.20) | | | -1.74*** (0.67) | -1.83** (0.71) | | |
| Mid OAR | 0.63*** (0.17) | 0.54*** (0.12) | | | -0.04 (0.16) | -0.05 (0.17) | | |
| High OAR | 0.55*** (0.17) | 0.52*** (0.18) | | | -0.41*** (0.15) | -0.41*** (0.15) | | |
| Low Rank | | | -0.19 (1.61) | 0.27 (1.56) | | | -10.70*** (3.56) | -11.43*** (3.91) |
| Mid Rank | | | 2.07*** (0.66) | 1.75*** (0.43) | | | 0.01 (0.33) | 0.00 (0.36) |
| High Rank | | | 1.42 (3.88) | 1.88 (3.24) | | | -5.66** (2.21) | -5.86** (2.40) |
| Constant | -0.53** (0.24) | -0.42* (0.22) | -0.61*** (0.22) | -0.60** (0.24) | -1.37** (0.63) | -1.42** (0.68) | 2.00*** (0.62) | 2.13*** (0.68) |
| Observations | 111318 | 111318 | 111328 | 111328 | 111318 | 111318 | 111328 | 111328 |
| Adjusted R^2 (%) | .02 | .02 | .02 | .02 | .05 | .05 | .03 | .03 |

Table 13: Twitter Activity and Future Fund Performance

Note: This table shows estimation results of regressions of fund's monthly alpha (in basis points) on the number of tweets, past performance, and control variables. Columns (1) to (5) use the monthly alpha using the CAPM, Fama and French 3 Factors, Carhart's, Fama and French 5 Factors, and Fama and French 5 Factors plus Momentum models. Retail is a dummy variable that takes the value of one if the fund is classified as a retail fund. Standard errors clustered at the family, month, and family-month levels are presented in parentheses. All variables are defined in Section (2). Finally ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level, respectively.

| | (1) $\alpha_{t \rightarrow t+1}$ bp | (2) $\alpha_{t \rightarrow t+1}^{FF3}$ bp | (3) $\alpha_{t \rightarrow t+1}^{FF3+M}$ bp | (4) $\alpha_{t \rightarrow t+1}^{FF5}$ bp | (5) $\alpha_{t \rightarrow t+1}^{FF5+M}$ bp |
|---|--|--|--|--|--|
| Number of Tweets (1,000) | 13.29 (9.98) | 12.42 (8.90) | 11.70 (8.49) | 10.85 (8.41) | 11.12 (8.00) |
| $\alpha_{t-3 \rightarrow t}$ bp | 0.00 (0.01) | | | | |
| $\alpha_{t-3 \rightarrow t}^{FF3}$ bp | | -0.01 (0.01) | | | |
| $\alpha_{t-3 \rightarrow t}^{FF3+M}$ bp | | | -0.01 (0.01) | | |
| $\alpha_{t-3 \rightarrow t}^{FF5}$ bp | | | | -0.00 (0.01) | |
| $\alpha_{t-3 \rightarrow t}^{FF5+M}$ bp | | | | | 0.00 (0.01) |
| log(TNA (USD millions)) | -8.97*** (0.99) | -10.56*** (0.79) | -10.00*** (0.75) | -7.55*** (0.74) | -7.44*** (0.71) |
| Expense Ratio | 307.34* (181.60) | 366.30* (190.62) | 291.48* (161.91) | 312.03* (162.79) | 274.91* (146.77) |
| Volatility (%) | -205.80 (199.07) | 40.06 (191.85) | 26.96 (182.80) | 207.31 (188.01) | 84.66 (179.32) |
| Family Age (log) | -6.49* (3.84) | -3.81 (3.34) | -3.83 (3.20) | -4.37 (3.15) | -4.23 (3.01) |
| Fund Age (log) | 17.10*** (3.38) | 12.20*** (2.79) | 12.95*** (2.67) | 9.95*** (2.61) | 11.14*** (2.48) |
| Family Size (log) | -5.16* (2.84) | -4.14* (2.23) | -2.93 (2.12) | -2.23 (2.08) | -0.84 (1.97) |
| Flow to Same Category (%) | 24.87*** (1.17) | 18.97*** (1.03) | 17.33*** (0.98) | 16.38*** (0.90) | 15.12*** (0.87) |
| Retail | 1.72 (3.00) | 2.47 (2.58) | 3.33 (2.45) | 1.60 (2.39) | 2.81 (2.28) |
| log(Number of Funds) | 6.53 (4.47) | 3.29 (3.74) | 1.82 (3.51) | 2.35 (3.48) | 1.02 (3.26) |
| Constant | 22.80 (35.23) | 28.63 (29.73) | 17.20 (28.35) | 7.71 (27.97) | -3.76 (26.49) |
| Observations | 511965 | 511965 | 511965 | 511965 | 511965 |
| Adjusted R^2 (%) | 9.6 | 7.04 | 6.97 | 6.88 | 7.04 |
| Time FE | Yes | Yes | Yes | Yes | Yes |
| Inv. Category FE | Yes | Yes | Yes | Yes | Yes |
| Fund FE | Yes | Yes | Yes | Yes | Yes |
| Fund Family FE | Yes | Yes | Yes | Yes | Yes |

Table 14: Twitter Activity, Fund Characteristics, and Flows

This table shows estimation results for regressions of mutual fund flows in percentage points on the number of tweets, control variables, and its interactions. Control variables are defined in Section (2). Standard errors clustered at the month, fund family, and month-fund family level, are shown in parentheses. All variables are defined in Section (2). Finally ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---------------------------|--------------------|---------------------|---------------------|--------------------|--------------------|--------------------|--------------------|
| Number of Tweets (1,000) | 1.08** (0.47) | 0.43 (0.27) | 1.11*** (0.30) | -0.08 (1.34) | -2.57*** (0.64) | -3.42* (1.98) | 0.57*** (0.17) |
| × log(TNA (USD millions)) | -0.10 (0.07) | | | | | | |
| × Volatility (%) | | 1.96 (5.89) | | | | | |
| × Expense Ratio | | | -63.33** (24.60) | | | | |
| × Fund Age (log) | | | | 0.11 (0.25) | | | |
| × Family Size (log) | | | | | 0.28*** (0.06) | | |
| × Family Age (log) | | | | | | 0.60** (0.30) | |
| × Past Flows | | | | | | | 0.09** (0.05) |
| log(TNA (USD millions)) | -0.63*** (0.03) | -0.64*** (0.03) | -0.64*** (0.03) | -0.64*** (0.03) | -0.64*** (0.03) | -0.64*** (0.03) | -0.64*** (0.03) |
| Low OAR | 0.33*** (0.03) | 0.33*** (0.03) | 0.33*** (0.03) | 0.33*** (0.03) | 0.33*** (0.03) | 0.33*** (0.03) | 0.33*** (0.03) |
| Mid OAR | 0.67*** (0.02) | 0.67*** (0.02) | 0.66*** (0.02) | 0.67*** (0.02) | 0.66*** (0.02) | 0.66*** (0.02) | 0.66*** (0.02) |
| High OAR | 0.79*** (0.04) | 0.79*** (0.04) | 0.79*** (0.04) | 0.79*** (0.04) | 0.79*** (0.04) | 0.79*** (0.04) | 0.79*** (0.04) |
| Expense Ratio | 5.08** (2.35) | 5.04** (2.34) | 18.41*** (6.15) | 5.03** (2.34) | 5.23** (2.36) | 5.14** (2.35) | 5.07** (2.34) |
| Volatility (%) | -9.87*** (1.14) | -10.03*** (1.27) | -9.87*** (1.14) | -9.86*** (1.14) | -9.81*** (1.14) | -9.83*** (1.14) | -9.89*** (1.14) |
| Family Age (log) | -0.20** (0.09) | -0.20** (0.09) | -0.20** (0.09) | -0.20** (0.09) | -0.19** (0.09) | -0.21** (0.09) | -0.20** (0.09) |
| Fund Age (log) | -1.90*** (0.11) | -1.89*** (0.10) | -1.90*** (0.10) | -1.89*** (0.11) | -1.89*** (0.10) | -1.89*** (0.10) | -1.90*** (0.10) |
| Family Size (log) | 0.27*** (0.05) | 0.27*** (0.05) | 0.28*** (0.05) | 0.27*** (0.05) | 0.26*** (0.05) | 0.27*** (0.05) | 0.27*** (0.05) |
| Past Flows | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) |
| Flow to Same Category (%) | 0.44*** (0.01) | 0.44*** (0.01) | 0.44*** (0.01) | 0.44*** (0.01) | 0.44*** (0.01) | 0.44*** (0.01) | 0.44*** (0.01) |
| Constant | 10.84*** (0.90) | 10.82*** (0.90) | 10.71*** (0.90) | 10.81*** (0.90) | 10.84*** (0.90) | 10.86*** (0.90) | 10.85*** (0.90) |
| Observations | 453572 | 453572 | 453572 | 453572 | 453572 | 453572 | 453572 |
| Adjusted R^2 (%) | 13.22 | 13.22 | 13.22 | 13.22 | 13.22 | 13.22 | 13.22 |
| Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Inv. Category FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Fund FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Fund Family FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

Table 15: Twitter Activity, Fund Characteristics, and Future Fund Performance

Note: This table shows estimation results of regressions of fund's monthly alpha (in %) on the number of tweets, past performance, control variables, and their interactions. Variables with (z) are standardized to have a mean of zero and a standard deviation of one. Retail is a dummy variable that takes the value of one if the fund is classified as a retail fund. Standard errors clustered at the family, month, and family-month levels are presented in parentheses. All variables are defined in Section (2). Finally ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level, respectively.

| | (1) | (2) | (3) | (4) |
|---------------------------------|---------------------|---------------------|---------------------|---------------------|
| Number of Tweets (1,000) | 7.72 (10.35) | 11.36 (9.14) | 14.67 (9.86) | 13.40 (9.95) |
| × Expense Ratio (z) | -24.47*** (6.83) | | | |
| × Family Size (log) (z) | | 4.77 (7.01) | | |
| × Family Age (log) (z) | | | -3.57 (9.06) | |
| × Past Flows (z) | | | | -1.11 (5.38) |
| Expense Ratio (z) | 6.71*** (1.89) | | | |
| $\alpha_{t-3 \rightarrow t}$ bp | 0.00 (0.01) | 0.00 (0.01) | 0.00 (0.01) | 0.00 (0.01) |
| log(TNA (USD millions)) | -8.98*** (0.99) | -8.99*** (0.99) | -8.96*** (1.00) | -8.86*** (0.99) |
| Volatility (%) | -206.30 (199.08) | -204.97 (199.08) | -206.21 (199.06) | -209.06 (199.12) |
| Family Age (log) | -6.71* (3.83) | -6.35* (3.84) | | -6.52* (3.84) |
| Fund Age (log) | 16.72*** (3.39) | 17.17*** (3.38) | 17.05*** (3.38) | 16.23*** (3.37) |
| Family Size (log) | -5.04* (2.84) | | -5.15* (2.84) | -5.15* (2.84) |
| Flow to Same Category (%) | 24.92*** (1.17) | 24.87*** (1.17) | 24.86*** (1.17) | 24.93*** (1.17) |
| Retail | 1.95 (3.01) | 1.61 (3.00) | 1.79 (2.98) | 1.75 (3.00) |
| log(Number of Funds) | 6.55 (4.47) | 6.53 (4.47) | 6.56 (4.47) | 6.40 (4.47) |
| Family Size (log) (z) | | -9.76* (5.24) | | |
| Expense Ratio | | 309.13* (181.78) | 306.48* (181.39) | 311.17* (182.56) |
| Family Age (log) (z) | | | -4.19* (2.48) | |
| Past Flows (z) | | | | -1.26*** (0.45) |
| Constant | 28.01 (35.22) | -31.93 (33.45) | -17.97 (30.76) | 27.20 (35.26) |
| Observations | 511965 | 511965 | 511965 | 511965 |
| Adjusted R^2 (%) | 9.61 | 9.6 | 9.6 | 9.61 |
| Time FE | Yes | Yes | Yes | Yes |
| Inv. Category FE | Yes | Yes | Yes | Yes |
| Fund FE | Yes | Yes | Yes | Yes |
| Fund Family FE | Yes | Yes | Yes | Yes |

Table 16: Twitter Activity, Investor Sophistication, and Flows

This table shows estimation results for regressions of mutual fund flows in percentage points on the number of tweets, control variables, and an interaction with a retail dummy. Control variables are defined in Section (2). Standard errors clustered at the month, fund family, and month-fund family level, are shown in parentheses. All variables are defined in Section (2). Finally $***$, $**$, and $*$ denote statistical significance at the 1%, 5%, and 10% significance level.

| | (1) | (2) |
|--|--------------------|--------------------|
| Number of Tweets (1,000) | 0.49*** (0.16) | -0.18 (0.24) |
| Retail | 0.18** (0.07) | 0.11 (0.07) |
| Number of Tweets (1,000) \times Retail | | 0.87*** (0.26) |
| Low OAR | 0.33*** (0.03) | 0.33*** (0.03) |
| Mid OAR | 0.67*** (0.02) | 0.67*** (0.02) |
| High OAR | 0.79*** (0.04) | 0.79*** (0.04) |
| log(TNA (USD millions)) | -0.64*** (0.03) | -0.64*** (0.03) |
| Expense Ratio | 5.04** (2.34) | 5.05** (2.34) |
| Volatility (%) | -9.82*** (1.14) | -9.77*** (1.14) |
| Family Age (log) | -0.18* (0.09) | -0.18* (0.09) |
| Fund Age (log) | -1.89*** (0.10) | -1.89*** (0.10) |
| Family Size (log) | 0.28*** (0.05) | 0.27*** (0.05) |
| Past Flows | 0.14*** (0.01) | 0.14*** (0.01) |
| Flow to Same Category (%) | 0.44*** (0.01) | 0.44*** (0.01) |
| Constant | 10.53*** (0.90) | 10.60*** (0.90) |
| Observations | 453572 | 453572 |
| Adjusted R^2 (%) | 13.22 | 13.22 |
| Time FE | Yes | Yes |
| Inv. Category FE | Yes | Yes |
| Fund FE | Yes | Yes |
| Fund Family FE | Yes | Yes |

Table 17: Twitter Activity, Investor Sentiment, and Flows

This table shows estimation results for regressions of mutual fund flows in percentage points on the number of tweets, control variables, and an interaction with the aggregate orthogonal sentiment variable of Baker and Wurgler (2006). Control variables are defined in Section (2). Standard errors clustered at the month, fund family, and month-fund family level, are shown in parentheses. All variables are defined in Section (2). Finally ***, **, and * denote statistical significance at the 1%, 5%, and 10% significance level.

| | (1) | (2) | (3) | (4) | (5) |
|------------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Sentiment (Orthogonalized) | 0.10 (0.11) | 0.04 (0.11) | 0.08 (0.11) | 0.08 (0.11) | 0.10 (0.11) |
| Retail | | | | 0.06 (0.07) | 0.09 (0.07) |
| Number of Tweets (1,000) | | 0.55 (0.41) | | -0.25 (0.52) | |
| × Sentiment (Orthogonalized) | | 1.26* (0.67) | | 1.18* (0.65) | |
| × Retail | | | | 0.92*** (0.35) | |
| Positive Tweets (1,000) | | | 0.70** (0.30) | | -1.08 (0.91) |
| × Sentiment (Orthogonalized) | | | 1.48* (0.83) | | 1.44* (0.86) |
| × Retail | | | | | 1.80* (0.99) |
| Low OAR | 0.35*** (0.06) | 0.35*** (0.06) | 0.35*** (0.06) | 0.35*** (0.06) | 0.35*** (0.06) |
| Mid OAR | 0.66*** (0.05) | 0.66*** (0.05) | 0.66*** (0.05) | 0.66*** (0.05) | 0.66*** (0.05) |
| High OAR | 0.78*** (0.11) | 0.78*** (0.11) | 0.78*** (0.11) | 0.78*** (0.11) | 0.78*** (0.11) |
| log(TNA (USD millions)) | -0.63*** (0.07) | -0.63*** (0.07) | -0.63*** (0.07) | -0.63*** (0.07) | -0.63*** (0.07) |
| Expense Ratio | 4.68 (4.90) | 4.61 (5.04) | 4.70 (4.99) | 5.17 (4.87) | 5.29 (4.83) |
| Volatility (%) | -5.91*** (1.47) | -5.84*** (1.40) | -5.89*** (1.46) | -5.74*** (1.40) | -5.70*** (1.46) |
| Family Age (log) | -0.22 (0.13) | -0.21 (0.13) | -0.21 (0.13) | -0.20 (0.13) | -0.20 (0.13) |
| Fund Age (log) | -1.92*** (0.16) | -1.93*** (0.16) | -1.93*** (0.16) | -2.02*** (0.16) | -2.04*** (0.16) |
| Family Size (log) | 0.30*** (0.10) | 0.30*** (0.10) | 0.30*** (0.10) | 0.28*** (0.11) | 0.28*** (0.11) |
| Past Flows | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) | 0.14*** (0.01) |
| Flow to Same Category (%) | 0.49*** (0.04) | 0.49*** (0.04) | 0.49*** (0.04) | 0.50*** (0.04) | 0.50*** (0.04) |
| Constant | 10.60*** (1.59) | 10.64*** (1.59) | 10.67*** (1.57) | 11.18*** (1.61) | 11.24*** (1.63) |
| Observations | 453572 | 453572 | 453572 | 453572 | 453572 |
| Adjusted R^2 (%) | 13.07 | 13.08 | 13.07 | 13.08 | 13.08 |
| Time FE | No | No | No | No | No |
| Inv. Category FE | Yes | Yes | Yes | Yes | Yes |
| Fund FE | Yes | Yes | Yes | Yes | Yes |
| Fund Family FE | Yes | Yes | Yes | Yes | Yes |