INSEAD The Business School for the World®

Working Paper

2024/23/FIN

The Early Bird Catches the Worm: How Lasting is the Value of New, Alternative Data?

Prince Elvis Asamoah
City University of Hong Kong, peasamoah2-c@my.cityu.edu.hk

Massimo Massa
INSEAD, massimo.massa@insead.edu

Albert Kwame Mensah
HEC Paris, mensah@hec.fr

Vicki Wei Tang
Georgetown University, wt29@georgetown.edu

This Draft: March 4, 2024

We investigate how the information content of alternative data is impounded in prices and the duration of its value to mutual fund managers. Using a regression discontinuity design, we document that mutual funds increase their loadings on specific stocks by 0.7%-3% in response to exogenous, rounding-induced 1-percentage-point increase in ratings from customer-generated comments about companies' products and services on social media platforms. This effect is more pronounced when information asymmetry is greater. Funds relying more on such data yield higher abnormal future returns and exhibit better stock-picking and market-timing abilities. This effect dissipates when the data becomes public.

Keywords: Artificial Intelligence; Big Data; Alternative Data; Newsanalytics; Stock Market Analysis; Trading Strategies

*We appreciate helpful comments from workshop participants at Georgetown University, especially those from Xiaofei Zhao, Jennie Bai, Lee Pinkowitz, Clare Liang, and Reining Petacchi.

"[We] have historically utilized a large set of company-specific data like publicly available financial statement, as well as market data like prices, returns, volumes, etc. With the growth and availability of non-traditional data sources such as internet web traffic, patent filings and satellite imagery, we have been using more nuanced and sometimes unconventional data to help us gain an informational advantage and make more informed investment decisions." (GSAM 2016).

1. Introduction

The rise of Big Data has been exponential: overall, it is estimated that over 2.5 quintillion bytes of data are created every single day, and 1.7MB of data are created every second for every person as of the year-end of 2020 (Domo 2021). This has impacted the financial markets, allowing quant traders and financial players to design new trading strategies and supposedly improve their performance. Among the big winners of the new data game are (and maybe mostly) the providers of the new data, as the abandonment of the paradigm of efficient markets has incentivized the creation and sale of new, "alternative" [1] datasets.

In addition to existing conventional data (such as accounting data and market data), mountains of *alternative* data have become increasingly available to the investing community, and innovations in information technology and the growth of online data flows have significantly expanded the set of alternative data upon which capital market participants make decisions (e.g., Liberti and Petersen 2019).[2] This vast amount of alternative information and the constant arrival of new data raises the question of the value of a new alternative dataset and for how long its usage

[1] Alternative data (also called nontraditional or unconventional data) is almost now a conventional term, at least, in finance, referring to hi-tech information sources such as satellite imagery, geolocation data, social media data, et cetera.

[2] FinTech is "changing the information available to investors, how it is delivered, and how it is converted into investing decisions" (Liberti and Petersen 2019, p.29) and big data is now more, than ever providing "more information for sophisticated players such as institutional investors" (Goldstein, Spatt and Ye 2021, p.3222).

warrants the cost of acquisition. This question of the value of information is at the core of finance and economics (e.g., Admati, 1985; Admati and Pfleiderer 1986, 1987, 1988, 1990). Indeed, unlike other products, information loses its value as more people use it, and its sellers may decide to restrict the access to the information (Admati and Pfleiderer 1995, 1996, 1997).

However, till now, we have no quantification of the real value of a "*priced," new source* of information and its lifespan. Indeed, the traditional focus has been on the impact on the market of *new releases* of conventional sources of information – i.e., analysts' forecasts, recommendations or revisions, earnings releases, management-issued earnings forecasts.[3] The traditional information producers in these settings provide new "hard" information every quarter,[4] but use the *same* fundamentals-related information (e.g., cash flow forecasts).

The media literature, instead, has showed the impact of a structurally new and different source of information. However, such information is not produced with the explicit goal to be sold in the financial markets. This makes it difficult to assess its real value, given the difficulty to quantify its value in the main market in which such information is produced and disseminated in (i.e., newspaper audience).

Overall, to the best of our knowledge, traditional information studies examine the value of information at every quarter, but do not investigate whether there is a structural change with *new, different data that is explicitly created to be sold in the financial markets and how long it lasts*.

[3] The same is the case for "new usages" of existing data. For example, McLean and Pontiff (2016), focusing on the post-publication prediction power of 97 standard available financial and accounting variables shown by the finance literature to predict cross-sectional stock returns, document that the very publication of their predicting power reduces the post-publication ability of the involved variables to predict stock returns and therefore their value. They show that portfolio returns are 26% lower out-of-sample and 58% lower post-publication, implying a 32% lower return from *publication-informed* trading.

[4] For example, in the analysts' forecast revision setting, Gleason and Lee (2003) show that a substantial portion of the delayed price adjustment following forecast revisions at time *t* occurs around *subsequent* earnings-announcement and *subsequent* forecast-revision dates, implying that it is difficult to cleanly track all the way, the entire value-life of a new piece of information previously released at time *t*.

That is, the focus of prior studies has never been a direct assessment of the entire "*life-value*" of a "*priced*," "*new source*" of information. The contribution of this paper is to address this largely unexplored issue.

We focus on the creation of a new dataset offered on the market ("LikeFolio"), and investigate whether investors use such information, the speed at which the information content of this data is impounded into the trading process, and the duration of the value of such data to its users. That is, while prior studies examined how the market reacts to the spread of news that was not created to be sold to (asset) managers,[5] this paper focuses on the "value life" of a dataset that was explicitly created to cater to decision makers in the financial markets – i.e., asset managers and corporate CEOs.

Unlike media, the new information is not just stirring uniformed traders' attention but has a clear informational content that depends on the number of investors who use it. While such information is expensive and likely to be purchased by professional "informed" traders, its value still decreases with the number of investors who use it.

This has a major econometric implication. While the key issue facing the media literature is whether media induces attention-driven reactions by investors that are not related to the (almost non-existent) information content, in our case, the role of information is critically linked to the fact that the information is valuable and not already contained in the stock prices. That is, the value is there only if the informational content is not already contained in the existing information outlets such as prices and media analytics – and will deteriorate as more investors' trades impound it in the prices. The focus on information allows us to steer clear of the need of the traditional literature to distinguish information from attention and the issues of spurious correlation.

[5] In the realm of the media literature, the focus has mostly been on the way media drives attention of unsophisticated, less informed investors and generates price reversals (e.g., Tetlock 2007).

In terms of the informational content, we assess its value by relying on a unique experiment that exploits randomness created by exogenous, pre-determined rules of rounding (e.g., Hollenbeck, Moorthy and Proserpio 2019; Huang 2020; Derrien, Garel, Petit-Romec and Weisskopf 2023). In particular, we focus on one important alternative data and primitive information source available to financial markets in large volumes: the information contained in aggregated opinions of individual customers following focal firms on social media and online marketing platforms (Tang 2018; Huang 2018).[6,7] There is abundant anecdotal evidence supporting the idea that investors (particularly sophisticated investors) use this alternative data. For instance, Greenwich Associates LLC—a company providing market intelligence and advisory services to the financial services industry—reports that approximately 30 percent of institutional investors surveyed say the purchase of information on social media activity and content influenced their investment decision (see Greenwich Associates' press release on April 16, 2015).[8,9] In another example, a CNBC article reports that the share price of Apple Inc. surged "in the nine months after Apple's 2017 event, when social media intent-to-purchase mentions soared for the iPhone 8 and X."[10] Goldstein *et al*. (2021) highlight the significant impact of social media during the GameStop episode in January 2021 which saw the hedge fund Melvin Capital lose 53%.

---

[6] Customers make and unmake firms, as they are the lifeblood of business, the origin and source of a firm's sales (Petusevsky 2010; Tang 2018; Walker 1990; Webster 1994). Thus, compared to information produced by other capital market participants (such as news media, analysts, rating agencies, etc.), information produced directly by customers themselves is a *primitive* information source.

[7] On the economic significance of this information source, Tang (2018) for instance found that a one standard-deviation increase in the number of customer tweets about past purchases or future purchase intentions is associated with a 6.4 percent increase in future sales growth and 0.71 percent in upcoming sales growth that is not anticipated by analysts.

[8] https://www.greenwich.com/press-release/social-media-influencing-investment-decisions-global-institutions (accessed March 11, 2023)

[9] This era of 'Big Data' and Artificial Intelligence (AI) has made accessible simple tools for processing and learning from volumes of social media posts (see, e.g., GASM [2016] for quantitative portfolio managers' perspectives).

[10] See https://www.cnbc.com/2018/09/14/interest-in-new-apple-products-lowest-since-2015-keynote-hinting-at-months-of-anemic-stock-returns.html (accessed March 8, 2023)

We focus on the dataset that aggregates information on social media created and sold by the data provider, LikeFolio. The following testimonials from certain asset managers report that they use LikeFolio data: "*I use LikeFolio to research the consumer-facing companies in our portfolio and discover fresh opportunities I hadn't even considered… Unlike other social-data providers, you provide significant alpha that isn't highly correlated to our other data sets*." [11,12] Likefolio aggregates consumer posts about products and brands on Twitter and displays customer satisfaction ratings for companies associated with these products and brands. One key feature of LikeFolio's ratings data is that the ratings, ranging on a 0-100 scale, are rounded to the nearest whole number when displayed on the app or investors' dashboard. This generates a clear set of cutoff points or thresholds (i.e., 0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, …, 97.5, 98.5, and 99.5). For instance, an unrounded score of 79.31 will be *rounded down* and displayed as 79 on the LikeFolio app or investor's dashboard, whereas an unrounded score of 79.77 will be *rounded up* and displayed as 80. Having access to the unrounded (exact) customer satisfaction scores, and their proximity to the cutoff points, we can identify rating pairs straddling the cutoff points – i.e., the ones just above as well as the ones just below the cutoff points, which would be ultimately rounded up versus down. Critically, the *assignment to groups just above or below these cutoff points is random* because it is determined by an exogenous "pre-determined rule of rounding". We can, therefore, assess the impact of random assignments of the customer ratings on investor trading and stock prices, and quantify the life value of such information.

We focus on the behavior of one specific set of rational and sophisticated investors: the mutual fund managers. By the year-end 2020, the US mutual fund industry recorded $23.9 trillion in total net assets, with equity funds alone making up 53 percent of US mutual fund net assets. In

[11] https://likefolio.com/quant-funds/ (accessed July 21, 2023).
[12] https://likefolio.com/hedge-fund-managers/ (accessed July 21, 2023).

2020, an estimated 102.5 million Americans in 58.7 million households owned mutual funds. Mutual funds had net inflows of $205 billion in 2020, equivalent to 1.0 percent of year-end 2019 net assets.[13] Despite ample anecdotal evidence, there is currently no systematic, large-sample *causal* evidence (to our knowledge) on how the mutual fund industry is leveraging alternative datasets in their investment strategies and whether these trades are profitable (e.g., Huang 2018).[14]

We proceed as follows. First, we investigate how and to what extent mutual funds' trading relies on the information contained in customer-based ratings. We exploit the above-mentioned random assignment and use a regression discontinuity design (RDD) to identify whether mutual fund trades are based on customer-based ratings about a company's products and services on social media platforms. As expected, we find a marked discontinuity in fund trades around cutoff points in customer satisfaction ratings. In particular, after controlling for the exact (unrounded) customer satisfaction scores on social media and stock and fund characteristics, the portfolio weight of firms with unrounded ratings above the discontinuity (rounded–up firms) is 0.7-3.0 percent higher than the weight of firms with unrounded ratings below the discontinuity (rounded-down firms). The difference is economically significant: the increase of 0.7-3.0 percent in portfolio weight in response to an exogenous, rounding-induced 1-point increase (on a scale of 0-100) in Twitter-based consumer satisfaction scores corresponds to 2.7%-11.5% of one standard deviation in the change in portfolio weights. Mutual funds rely more on social media signals when information asymmetry is more pronounced as proxied by greater analysts' forecast errors, more volatile cash flows, and poorer accounting quality.

[13] https://www.icifactbook.org/21_fb_ch3.html

[14] For example, Huang (2018) find that customer-based wisdom of crowds, while positively associated with stock pricing, is positively associated with net buys by hedge funds, but not by non-hedge funds.

Next, we investigate whether mutual funds profit from trades relying on customer-based ratings. We identify those funds that rely more heavily on customer-based ratings using the same methodology of Kacperczyk and Seru (2007) based on the R^2 of the regression of percentage changes in fund managers' portfolio holdings on changes in *unrounded (exact)* customer satisfaction scores.[15] We document that, after controlling for funds' reliance on public information (such as analyst recommendations), funds with a greater reliance on LikeFolio-provided customer-based ratings yield higher abnormal performance in the next quarter. In particular, the four-factor Carhart alpha subsequently increases by 1.21% (0.038*0.32) in response to a one-standard-deviation increase in funds' reliance on LikeFolio-provided customer-based ratings. Tests based on alternative measures of reliance on LikeFolio-provided customer satisfaction scores deliver similar results both qualitatively and quantitatively. Overall, the second set of findings suggest that mutual funds with a greater reliance on alternative data earn abnormally high returns in the future. When we identify the sources of mutual fund outperformance, we find that such fund outperformance reflects both better stock picking and better market timing.

After establishing that customer-based ratings impact mutual fund trades and positively impact their performance, we next investigate the lifespan of the value of the alternative data. That is, we investigate whether the value of such data declines as more fund managers adopt it. We gather information on major publicity events of the data provider and identify the period from the third quarter of 2013 to the first quarter of 2014 as the publicity period when customer satisfaction scores gathered from social media gains public visibility. We find that mutual funds relying more on LikeFolio-provided customer-based ratings earn higher abnormal returns *only before* the

[15] We use this method to identify asset managers relying on LikeFolio data because, for reasons of client confidentiality, LikeFolio does not disclose its list of clients publicly or privately (upon request).

publicity period. That is, the value is high only when few investors use such information. However, the abnormal returns associated with the usage of the data provider's customer-based ratings dissipate *during* and *after* the publicity period, presumably because more and more investors are aware of such data after the publicity. This is consistent with the intuition that as more and more investors trade on such data, its information content is incorporated into the stock price in a more timely and efficient manner. Therefore, the abnormal fund returns disappear or decay when the data becomes public, and thus mutual funds can no longer outperform based on social media signals.

We structure the remainder of the paper as follows. Section 2 reviews key literature and discusses our contributions to the literature. Section 3 describes the data sources and our research methodology. Section 4 discusses the results on the impact of the information content of LikeFolio data on mutual fund portfolio choice. Section 5 discusses the results on the performance implications of relying on the LikeFolio data. Section 6 examines the lifespan of the value of LikeFolio data. Section 6 then concludes this paper.

2. Literature Review

This paper is related to several major strands of research. The first is the literature on information collection and processing by mutual fund managers. This information may be private or public. Private information may originate in many ways: social or business connections to portfolio firms' management or board members (Cohen, Frazzini and Malloy 2008; Duan, Hotchkiss and Jiao 2018; Gu, Li, Yang and Li 2019), geographic proximity to portfolio firms (Coval and Moskowitz 2001), private interactions with firm management (Bushee, Jung and Miller 2011, 2017; Solomon and Soltes 2015). Public information may be priced as the one coming from analyst reports

(Kacperczyk and Seru 2007), changes in sell-side analysts' earnings forecasts and recommendations (Brown, Wei and Wermers 2014; Franck and Kerl 2013), or from sources such as media coverage (Engelberg and Parsons 2011; Peress 2014; Solomon, Soltes and Sosyura 2014; Kaniel and Parham 2017), earnings announcements (Baker, Litov, Wachter and Wurgler 2010), accounting accruals (Ali, Chen, Yao and Yu 2008). We contribute to this literature by zooming in on a key "hard" source of information that, unlike the standard analyst-based one, is not yet diffused and, therefore, presumably with high financial value. We contribute to this literature by following the value of such information over its life cycle. In doing this, we relate to the Folk theorem that trades are expected to be profitable only if the data is not yet used by the overall market – i.e., fund managers are quick to spot new information. Indeed, as a quantitative portfolio manager at GSAM, stated, "active management has always been about uncovering opportunities before they are priced in by the broader market... Today, we are able to process more data more quickly, to uncover insights and connections that aren't as obvious to other investors." (GSAM 2016). This is also consistent with the scarce value of an existing source of priced, but publicly used information such as analyst forecasts.

Second, we contribute to the literature on big data and stock markets. The literature has focused on the market reaction to news (e.g., Riordan, Storkenmaier, Wagener, and Zhang 2013; Gross-Klugmann and Hautsch 2011; Sinha 2012; Zhang 2013) or on the impounding of new information in the market (e.g., Foucault, Hombert, and Rosu 2013). This study contributes to this strand of literature by showing how priced information can lose its value as it gets diffused in the market and more investors use it .

Third, we relate to the literature on the importance of the media. The literature has mostly focused on the causal impact of media on financial markets and has investigated whether the media has an impact on the market that is distinct from the underlying information it conveys (e.g., Engelberg and Parsons 2011; Peress 2013). The focus has thus been more on attention than information. We diverge drastically as we directly focus on information and investigate its value as more investors use it.

Finally, we relate to the literature on the role of online marketing/advertising platforms and product review sites. Prior studies have mostly focused either on the consequences of investors' use of online media (e.g., Kelly and Tetlock 2013; Bartov, Faurel, and Mohanram 2018; Lee, Ma and Wang 2015; Chen, De, Hu and Hwang 2014; Bollen, Mao, and Zheng 2011; Curtis, Richardson, and Schmardebeck 2016; Mao, Wei, Wang, and Liu 2012; Sprenger, Tumasjan, Sandner, and Welpe 2013) or on the consequences of corporate use of online media (e.g., Blankespoor, Miller, and White 2014; Lee, Hutton, and Shu 2015; Jung, Naughton, Tahoun, and Wang 2018; Crowley, Huang, and Lu 2018). As such, the primary concern of this research stream has been about the effects of online media on firm-level and stock performance, but not on the financial market consequences of investors' use of online media. Two recent studies have merged to fill the gap in the literature (e.g., Huang 2018; Derrien, Garel, Petit-Romec and Weisskopf 2020). We contribute to this emerging area of research by providing the first, direct causal evidence on whether, how, and to what extent the priced public information disseminated on social data is used in mutual fund managers' trading activities.

3. Research Design

3.1 Data and sample

Our sample dataset merges data across five databases, namely, the Center for Research in Security Prices (CRSP) Mutual Fund Survivorship-bias-free database, the Thomson Reuters Institutional

Holdings database, the Compustat database, the CRSP stock price database and LikeFolio. We retrieve data on fund returns and other fund characteristics from the CRSP Mutual Fund database. We eliminate observations before the reported fund inception date on CRSP, funds whose total net assets are less than $10 million, and funds with missing names. For a fund to be included in our analysis, it must also have at least ten stocks in the portfolio holding.

We obtain data on fund holdings from the Thomson Reuters Institutional Holdings database, which collects and compiles institutional investors' portfolio disclosures filed with the Securities and Exchange Commission (SEC).[16] Following the methodology in Kacperczyk, Sialm, and Seru (2007), we match mutual funds to portfolio holdings information. We obtain stock price and return data from the CRSP stock files, and obtain accounting data from Compustat.

To examine whether mutual fund managers use alternative data to evaluate the prospects of portfolio firms, we analyze Twitter-based consumer satisfaction ratings of the portfolio firms' products and brands. The customer rating data come from LikeFolio, a professional data analytics outfit that sells "data and insights to professional investors, corporate research teams, and software providers" (https://likefolio.com/terms-of-use/).[17] LikeFolio has also developed specific tools that can be integrated into the research workflow of hedge funds, quant funds, platform managers, and individual investors. Further, LikeFolio supplies its products and services to decades-old and top-ranked professional brokers. An example is TD Ameritrade, which displays LikeFolio social data

[16] SEC mandates the filing of all equity positions more than either 10,000 shares or $200,000 in market value for all institutions with over $100 million under discretionary management. See, Parida and Teo (2018)

[17] In our setting, we do not directly observe which specific corporate research teams are clients of LikeFolio. (For confidentiality purposes, the data provider does not provide its list of clients.) Thus, as in a preponderance of empirical studies on decision-making in capital markets, we learn from the data and allow our RDD setup to pick up any random effect that could be interpreted as the risk analytics departments of *some* funds likely using the LikeFolio data in their risk analytics. This assumption is not far-fetched, considering the abundance of anecdotes earlier offered on asset managers using social data.

and insights on the dashboard of the broker's users' trading apps.[18] The data provider also offers a free app on the Apple App Store and Google Play that any member of the financial public can freely use to research specific stocks (see Figure 1A). At the end of 2022, for instance, LikeFolio (Twitter handle: @LikeFolio) had over 7,000 direct followers on Twitter. The above suggests that the data provider's social data and insights are both directly and indirectly known to some sections of the public.

During our sample period of 2012 through 2015, LikeFolio data were first privately offered to a few professional asset managers and corporates, and then later publicized by popular finance news media (see, e.g., articles in CNN Money, Fox Business, Yahoo! Finance).[19,20] Thus, the heightened visibility associated with LikeFolio during this period ensures that capital market participants (including mutual fund managers) likely paid attention to LikeFolio and its products and services (including their consumer satisfactory ratings). Figures 1A (1B) show how free (purchased) LikeFolio data are directly displayed to data users.

To ensure the credibility of its social media data and insights, LikeFolio first uses a proprietary bot detection algorithm to sweep out fake social media posts. Using proprietary data and methods, LikeFolio develops a brand/product map for a list of public companies it covers and tracks total mentions, purchase intent mentions, and consumer happiness (i.e., satisfaction) for these brands on Twitter.[21] LikeFolio rates consumer satisfaction on a 0-100 scale and the ratings

---

[18] On their website, TD Ameritrade note that, "the social data discovery, filtering and analysis are provided by SwanPowers, LLC's *LikeFolio*. SwanPowers and TD Ameritrade are separate and unaffiliated firms." (https://www.tdameritrade.com/tools-and-platforms.html). The TD Ameritrade trading app has been recently ranked as the "Best App Overall" in 2023 (https://www.stockbrokers.com/guides/mobile-trading)

[19] CNN Money for instance ranked the LikeFolio app back in Dec. 2013 (part of our sample period 2012-2015) as a top financial app: https://money.cnn.com/gallery/pf/2013/12/11/best-financial-apps/2.html.

[20] Mutual fund managers that did not directly access the extensive data and insights across all LikeFolio-covered firms from either the LikeFolio dashboard or the dashboard of the TD Ameritrade app, can alternatively access free data and insights (in a limited fashion) only for specific firm(s) they follow (see Figure 1A).

[21] For a limited list of public firms covered by LikeFolio, see https://likefolio.com/resources/likefolio-coverage-list/. LikeFolio's scoring methodology is described in https://home.likefolio.com/knowledge-base/data-scoring-overview/

displayed on its app or investors' dashboard are rounded to the nearest whole number. By construction, there are 100 different rounding thresholds between 0 and 100. These rule-of-thumb rounding thresholds (that each unrounded, exact average rating score is close to) are 0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, …, 97.5, 98.5, and 99.5.[22] Relying on unrounded, exact average ratings close to these cutoff points, we can identify rating pairs just above versus just below these cutoff points, which would be ultimately rounded up versus down, respectively. For example, an unrounded average score of 99.49% will be rounded down and displayed as 99%, while an unrounded average score of 99.51% will be rounded up and displayed as 100%. Arguably, this assignment to groups just above or just below these cutoff points is random, because it is determined by an exogenous "pre-determined rule of rounding". To allow for documenting average estimates induced by rounding, we follow existing RDD and bunching studies (e.g., Best, Cloyne, Ilzetzki, and Kleven, 2020; Hollenbeck, Moorthy, and Proserpio 2019; Huang 2020; Derrien, Garel, Petit-Romec,and Weisskopf 2023) and pool together all the individual cutoff points into a single average cutoff. In subsequent analysis, we will estimate the causal effect of this random, discrete variation in consumer satisfaction ratings on our variable of interest, that is, mutual fund trades.

We initiate our sample construction by obtaining the LikeFolio consumer satisfaction ratings data from Tang (2018, Table 1). The sample covers 1,088 Compustat firms with 10,668 firm-year-quarter observations spanning the first quarter of 2012 (i.e., 2012Q1) and the last quarter of 2015 (i.e., 2015Q4). We then merge these consumer satisfaction rating data with Compustat fundamental data.

[22] They are the rounding thresholds that LikeFolio uses to round customer satisfaction scores to the nearest whole number (see, e.g., scores displayed on their app or the methodology section of their website).

We arrive at our final stock-level sample of 2,509 firm-year-quarter observations (number of firms or stocks = 727) by retaining only observations just above and just below the 100 evenly-spaced rounding thresholds (i.e., 0.5, 1.5, 2.5, 3.5, 4.5, 5.5, ..., 98.5, and 99.5) associated with LikeFolio consumer satisfaction ratings. Essentially, we retain only firm-year-quarter observations with unrounded customer satisfaction scores, which, when rounded to the nearest 0.5, take on the value of $n + 0.5$ (where n is a non-negative integer ranging from 0 to 99). Rounding thresholds are thus 0.5-based fractional values between pairs of successive non-negative integers: that is, thresholds are 0.5, 1.5, 2.5, ..., 98.5, and 99.5. While we lose 76.67 percent of firm-year-quarter observations (i.e., [10,668-2,509]/10,668) by applying this filter, our focus on observations subject to rounding will increase the power of the tests and isolate causal effects. We note that the sample attrition rate is similar in magnitude to that in prior studies. For instance, the constructed RDD sample is only 39% of the initial sample in Derrien et al. (2023) and 24% in Huang (2020). In the final RDD sample, all observations have a consumer satisfaction rating within 0.25 points of rounding thresholds; ±0.25 is, therefore, the maximum bandwidth in the final stock-level sample.

Because our research investigation focuses on mutual fund managers' trading behavior, we merge the above final stock-level sample of 2,509 observations with mutual fund trading data, arriving at our final, RDD-based fund-level sample of 592,889 fund-stock-year-quarter observations.[23] This fund-level sample is extremely large because, unlike the stock-level sample, where there is only one observation per stock-year-quarter, the fund-level sample has multitudes of observations per stock-year-quarter; essentially, multiple funds tend to hold multiple stocks in a year-quarter.

[23] While the RDD-based fund-level sample comprises of 592,889 fund-stock-year-quarter observations, the initial (pre-RDD) fund-level sample is composed of 1,327,439 fund-stock-year-quarter observations. This implies that the application of rounding rules results in a sample attrition rate of 55% for fund-level tests.

In Table 1, Panel A, we present the summary statistics of fund trade for all the funds (4,051 unique funds) in the CRSP Mutual Fund Database after the initial filtering described above and also present the fund trade statistics for only actively-managed equity mutual funds (1,802 unique funds). Our empirical analysis begins with all funds in the database and then we later focus on actively-managed equity funds (in contrast, we make use of passive funds as a placebo). Table 1, Panel B also presents the firm-level summary statistics for the original customer sentiment measure, rounding thresholds and the RDD indicator (i.e., ABOVE_VS_BELOW).

3.2 Empirical model

To estimate the causal impact of consumer purchase tweets on the trading behavior of fund managers, we adopt a regression discontinuity design (RDD) approach. Specifically, we exploit the feature of most consumer satisfaction ratings whereby third party-generated ratings disclosed to the public are the result of rounding upward or downward the exact, unrounded rating scores (see, e.g., Luca 2016; Hollenbeck, Moorthy and Proserpio 2019; Derrien, Garel, Petit-Romec and Weisskopf 2023; Huang 2020). In our setting, LikeFolio rounds down or up the average customer satisfaction scores to the nearest whole number (in percentages), which is displayed to the public. The fact that the assignment to groups just above or just below the rounding thresholds is based on an exogenous, pre-determined rounding rule rather than some firm-specific characteristics, allows us to use RDD to exploit the random, discrete variation in consumer satisfaction ratings around discontinuities created by rounding and estimate the causal effect of this variation on mutual fund trades. The idea is to attribute any discontinuous jump in trading behavior at these cutoff points to the discontinuous jump in customer satisfaction ratings.

As in other related studies using the RDD methodology, our RDD setup hinges critically on the following assumptions. First, asset manager-portfolio firm pairs that are close to the rounding thresholds do not choose to be above or below the said thresholds, implying that assignment to RDD groups is random. As our identification strategy depends critically on this assumption of randomness, we will formally test this assumption later in the paper. Second, capital market participants (including mutual fund managers) may possibly have access to both *rounded*, third-party-supplied rating scores and *unrounded* rating scores (that can be obtained either privately from external data analytic firms or generated internally by their own data analytics staff).[24,25] While we do not directly observe which form of these scores is used by fund managers to make portfolio allocation decision, an RDD is able to pick up the random effect of rounding insofar as decision makers use, to some degree, third-party-supplied rounded scores. This assumption does not therefore rule out the possibility that fund managers may also use exact, unrounded scores in their decision-making. This concern introduces a conservative bias against finding any RDD effect, implying that any documented effect in this setting would be a conservative estimate of the true effects. Third, as we do not know the exact dates when portfolio re-balancing decisions are made by each fund manager but only observe their respective quarterly trades at each firm-year-quarter-end, we assume that fund managers make decisions based on the rounding of the average daily ratings on the customers' products in a quarter.[26, 27] Nevertheless, in supplementary tests in the online appendix (see Table OA1) that exploits CRSP data on daily

[24] While external data analytic firms publicly display rounded rating scores, they may privately deliver the unrounded rating scores to their clients via data sale. In contrast, online marketing platforms (e.g., Amazon.com) *only* publicly display rounded scores (they do not provide unrounded scores).

[25] In the presence of data supply by a data analytics house or online marketing platform, it is possible that internal data analytics staffs may use their expertise to compute exact, unrounded scores, allowing for either *verifying* or *complementing* third party-supplied scores.

[26] This assumption is not only limited to our setting, but it seems to be the common assumption in almost all empirical equity and debt market studies employing cross-sectional pooled regressions to examine the information content of disclosures or fundamental data in a non-event study framework.

[27] In Figure 1B, we show that, using the LikeFolio dashboard, a user (e.g., fund manager) can easily obtain a rounded, 90-day moving average consumer satisfaction score at the end of a firm-year-quarter, which matches very well with the firm-year-quarter-end value of mutual fund trades.

aggregate trades, we, as a confirmatory check, use the daily setting (rather than the quarterly setting) to estimate the sensitivity of daily trading volume to daily consumer satisfaction scores.

Exploiting the quarterly setup, we specify the linear RDD model below as our primary model. In later specifications, as a robustness check, we also consider quadratic and cubic RDD models. We estimate:

$$
\begin{aligned}
&FUND\_TRADE_{f,i,q} \\
&\qquad = \beta_o + \beta_1\, ABOVE\ VS\ BELOW_{i,q} + \beta_2\, POSITIVE(unrounded)_{i,q} \\
&\qquad + \beta_{3\text{-}\infty}\, Firm\ Characteristics_{i,q} + FEs + \varepsilon_{f,i,q}
\end{aligned}
$$

-------- (1)

In Equation (1), *f*, *i* and *q* represent the mutual fund, portfolio firm (i.e., stock), and reporting year-quarter, respectively. $FUND\_TRADE_{f,i,q}$ measures the change in the weight (w) of stock *i* from quarter $q-1$ to quarter q in fund manager *f's* portfolio (Pinnuck 2003, p.814). That is, $FUND\_TRADE_{\ f,i,q} = w_{f,i,q} - w^{pq}_{f,i,q-1}$, where $w_{f,i,q} = \frac{P_{i,q}\, H_{f,i,q}}{\sum_{i=1}^{N}(P_{i,q}\, H_{f,i,q})}$; $w^{pq}_{f,i,q-1} = \frac{P_{i,q}\, H_{f,i,q-1}}{\sum_{i=1}^{N}(P_{i,q}\, H_{f,i,q-1})}$; $P_{i,q}$=price of stock *i* at quarter *q*; $H_{f,i,q}$=split-adjusted number of shares of stock *i* held by fund *f* at quarter *q*; and N= the number of different stocks held by each fund manager.[28]

$ABOVE\ VS\ BELOW_{i,q}$ is a dummy variable equal to 1 (0) if, in a given firm-year-quarter, an observation is just above (below) rounding thresholds. $POSITIVE(unrounded)_{i,q}$ is LikeFolio's '0-100 scale' customer satisfaction score, measured as one hundred (100) times "the ratio of the number of tweets that convey a positive assessment of products and brands over the number of tweets that convey a non-neutral (either positive or negative) assessment of products and brands" (Tang 2018, p.991). To facilitate the interpretation of the results, in all tabulations, we

[28] We winsorize the fund trade measure at the 1% and 99% levels.

provide estimates in which we re-scale this variable so that it uses the same 0-1 scale as the RDD indicator *ABOVE_VS_BELOW*.

$Firm\ Characteristics_{i,q}$ is a vector of proxies of characteristics of portfolio firms that may affect how fund managers allocate weights to specific stocks in their portfolios. The portfolio firm characteristics are firm size (*SIZE_ASSETS*), profitability (*PROFITABILITY, LOSS*), asset tangibility (*TANGIBILITY_NET*), indebtedness to creditors (*LEVERAGE*), growth prospects (*MARKET_TO_BOOK*, *SALES_GROWTH*), financial health (*Z_SCORE*), ready-to-sell inventories on hand (*INVENTORY*), length of operating cycle (*OPERATING_CYCLE*), cash-in-hand (*CASH_HOLDINGS*), and distributions to shareholders (*DIVIDEND*).

FEs is a vector of fixed effects at the time and fund levels, allowing us to control for heterogeneities in the panel data. Specifically, Time FEs allow us to control for the potential confounding influence of macro-level influences and shocks (e.g., economic growth, financial crisis, etc.) on portfolio choice. Fund FEs, on the other hand, control for the confounding effect of time-invariant fund-specific factors on portfolio choice. We do not impose a stock FE or stock's industry FE structure because the outcome variable (i.e., *FUND_TRADE*) is a differenced variable computed as the change in the holdings of (i.e., weight in) a stock by a fund from quarter q-1 to quarter q; as such, any potential confounding, time-invariant stock- or industry-level influence would have been directly purged via differencing (see, e.g., Roberts and Whited 2013, p.559).[29] We cluster the standard errors at the stock level to control for 3 types of correlations among fund observations: cross-sectional correlation among funds covering the same stock in the same period; the cross-sectional correlation among funds covering the same stock in different periods; and the

[29] "If the dependent variable is a first differenced variable, such as investment or the change in corporate cash balances, and if the fixed effect is related to the level of the dependent variable, then the fixed effect has already been differenced out of the regression, and using a fixed-effects specification reduces efficiency." (Roberts and Whited 2013, p.559)

time-series correlation of the same fund covering the stock in different periods. In untabulated results (available upon request), we alternatively use a two-way clustering method (i.e., fund + stock clustering), finding qualitatively similar results.

3.3 Validation of the Empirical model

We now assess whether the key identifying condition for the use of RDD has been met – i.e., for the observations close to the rounding thresholds, firm-level control variables should not systematically predict random assignment to groups (i.e., just above versus just below). To test this necessary condition, we use both univariate and multivariate analyses. For the univariate analysis, we test the null hypothesis that the difference in means of all firm characteristics between the treated group (i.e., rounded up group $ABOVE\_VS\_BELOW_{i,q} = 1$) and the control group (i.e., rounded down group $ABOVE\_VS\_BELOW_{i,q} = 0$) is not different from zero. As reported in Panel B of Table 1, the two-tailed t-test for the difference in means between the two groups fails to reject the null hypothesis across all twelve (12) firm characteristics. Specifically, the p-value for each variable is greater than 0.10, indicating that the difference in means between the two RDD assignment groups is not statistically significant.

In the multivariate analysis (unreported for brevity, but available upon request), we regress $ABOVE\_VS\_BELOW_{i,q}$ on all firm-level characteristics, controlling or not controlling for industry and time fixed effects.[30] The results show that almost none of the firm characteristics affect the assignment to the RDD groups, whether we use fixed effects or not. For instance, the p-value on all the twelve coefficient estimates is greater than 0.10 in the estimation based on the fund-level sample. Both the univariate and multivariate results support the assumption of randomization and the local exogeneity around rounding thresholds.

[30] In an alternative setup that uses the fund-level sample, we additionally control for fund fixed effects.

Next, we perform an additional test to further validate our assumption regarding the randomness of assignment to RDD groups. Specifically, we test if there is a manipulation of ratings close to the cut-off points (i.e., thresholds). This test aims to ensure a continuity of the density of the "running variable" $RATING\_RELATIVE\_TO\_THRESHOLD_{i,q}$ around the rounding thresholds. Essentially, it checks to ensure that there is no significant cluster of observations just below cut-off points (in the case of upward manipulation) and no clump of observations just above cut-off points (in the case of downward manipulation). To perform this test, we employ the statistical test of local polynomial density estimation (Cattaneo, Jansson and Ma, 2020, 2021a, 2021b). In untabulated results (based on the non-duplicative stock-level sample of 2,509 firm-year-quarter observations), we do not reject the null hypothesis of non-manipulation at cut-off points (i.e., t = -0.08, p-value = 0.94). This further corroborates our assumption that the assignment of observations to the RDD groups is random.

4. Information and Mutual Fund Portfolio Choice

4.1 Baseline Results

In this subsection, we formally examine the effect of exogenous, rounding-induced increases in Twitter-based consumer satisfaction scores on the weight that fund managers allocate to specific stocks in their portfolios. Specifically, we employ both graphical and regression analysis under the RDD setup to estimate the causal effects of consumer tweets on fund trades.

We begin with a graphical analysis to examine *visually* whether there is a discontinuous jump in fund trades at cutoff points, suggesting a causal effect of consumer satisfaction ratings on mutual fund portfolio allocations. For this analysis, we want to be close to cutoff points as much as possible, as discontinuity (if any) would occur at such cutoff points. Therefore, we not only focus on all observations that are within the maximum bandwidth of ±0.25 rating points of each rounding threshold (see Figure 2), but we also narrow down the bandwidth to ±0.15, ±0.10, and ±0.05 in alternative graphical presentations (see Figure 2).[31]

We then divide the consumer satisfaction rating space into an even number of bins on the left and right sides of cutoff points. We start with the bins at the pooled cutoff point of 0 (i.e., the position of any of the cutoff points), and gradually work our way out to the right and left of 0 to ensure that no bin contains both observations just above (i.e., the treatment group) and observations just below (i.e., the control group) cutoff points. For each bin, we compute the average trade in particular stocks and the corresponding midpoint value of $POSITIVE(unrounded)_{i,q}$ and plot average mutual fund trades for each bin on the vertical axis against the corresponding midpoint $POSITIVE(unrounded)_{i,q}$ value at each bin on the horizontal axis. We then use a linear fit to plot the mutual fund trades-consumer tweets relation.[32]

Using the maximum (different) bandwidth(s), Figure 2 plots the average fund trade (proxied by $FUND\_TRADE_{f,i,q}$) as a function of LikeFolio customer satisfaction ratings around cutoff points. We start with the visual representation based on the maximum bandwidth of ±0.25. The evidence presented in Figure 2 shows that based on a bin size of 40, as well as an optimal bin size (unreported for brevity, but available upon request), [33, 34, 35] there is a discontinuous (upward) jump in mutual fund trades at cutoff points. In the same Figure 2, we plot the fund trades-consumer satisfaction nexus for the narrow bandwidths ±0.15, ±0.10,and ±0.05. As displayed, we

[31] We re-center all observations around their respective cutoff points to 0 to allow for easily identifying the position of each observation relative to any of the 100 rounding thresholds.

[32] In alternative graphs (unplotted for brevity, but available upon request), using high-order polynomials such as quadratic fit yields qualitatively similar visual evidence as the linear fit.

[33] Adopting a bin size of 40 implies that we adopt 40 bins on left-hand side and another 40 bins on the right-hand side of rounding thresholds.

[34] Alternatively, adopting smaller or larger bin sizes produces qualitatively about the same visual evidence displayed in the current reported plots.

[35] Optimal bin size is determined via minimizing the integrated mean square error (IMSE). As suggested by Calonico, Cattaneo, and Titiunik (2015), IMSE-optimal binning allows for providing a plot that is well suited to detecting potential discontinuities in the underlying regression functions.

consistently see across all plots, qualitatively about the same discontinuous (upward) jump in fund trades at the cutoff points that we had observed under the maximum bandwidth.[36] The visual evidence suggests an increase in the weight that fund managers allocate to specific stocks in their portfolio due to rounding up versus rounding down, stocks' consumer satisfaction scores.

A concern with the visual evidence presented for the maximum and larger bandwidths in particular is the fact that the slope of the plot seems to be different below vs. above the pooled cutoff point. To address this concern, in the robustness checks, we additionally control for the slope of the RDD plot in regression estimations (see Table 5). The results do not differ.

We next provide a multivariate analysis that quantifies the precise economic magnitude of the effect of consumer satisfaction ratings on fund managers' allocation of weights to portfolio stocks. To this end, we utilize a multivariate estimation that allows us to control for the potential confounding effect of portfolio firm attributes, as well as control for heterogeneities across time and funds. In Table 2, we present the results of our linear RDD estimation using, first, the maximum bandwidth of ±0.25, and then, next, using the narrower bandwidths of ±0.15, ±0.10, and ±0.05.

We begin with the specifications involving the maximum bandwidth. In Panel A of Table 2, we find the coefficient on the RDD indicator to be positive (coeff. ranges between 0.007 and 0.008) and significant at the 10% level. The stability of coefficient across all 4 columns with vs. without portfolio firm-level controls and with vs. without FEs provides comfort that firm-level characteristics and therefore omitted variables in general are unlikely confounds to our results (see,

[36] We confirm that, in alternative tests (unreported for brevity, but available upon request), we continue to observe this behavior if we separate "sell" from "buy" and fixate only on increase in "buy" portfolio weight in stocks.

e.g., Altonji, Elder and Taber, 2005).[37, 38] Because the independent variable is an indicator and the dependent variable a proportion/weight (i.e., a percentage), we directly interpret the slope coefficient of 0.007-0.008 as an increase in net buys by 0.7-0.8 percent in response to an exogenous, rounding-induced 1-point increase (on a scale of 0-100) in Twitter-based consumer satisfaction scores.[39] Stated differently, portfolio firms with unrounded ratings above the discontinuity (i.e., those with corresponding displayed ratings that were rounded up) have a net buy portfolio allocation weight that is 0.007-0.008 higher than that of firms with unrounded ratings below the discontinuity (i.e., those with corresponding displayed ratings that were rounded down). We note that this increase in net buys by 0.7-0.8 percent corresponds to 2.7%-3% of one standard deviation in the change in portfolio weights, which is economically significant.

Next, we move to estimations involving narrower bandwidths in Panels B-D of Table 2. We document that both the economic magnitude and statistical significance increase as we narrow down the bandwidth. That is, the coefficient range first increases to 0.010-0.011 (p-values ranging between <0.05 and <0.10) for bandwidth equal to ±0.15. Then, when we further restrict the bandwidth to ±0.10 (±0.05), the coefficient range further increases to 0.019-0.022 (0.022-0.030), with p-values ranging between <0.01 and <0.05. Considering that economic magnitudes under smaller bandwidths are larger, we note that the full range of effect size is therefore an increase in

---

[37] Altonji et al. (2005) argue that the bias from the exclusion of observed controls is informative about the bias from the omission of unobserved controls (i.e,. in the presence of omitted variable bias, the coefficient on a variable of interest differs significantly wildly between specifications with and without controls).

[38] While the main specification adopts a "Fund + Time" FE structure, in untabulated tests (available upon request), we instead adopt a "Fund × Time" FE structure with vs. without control variables to control for the potential confounding influence of all *time-varying fund characteristics* (whether observed or unobserved) and we document about the same coefficient (i.e., coeff. = 0.007), which is also statistically significant at the 10% level. The stability of this coefficient across columns here again provides comfort that omitted fund characteristics are unlikely confounds to our results.

[39] We do not scale the coefficient by the mean value of *FUND_TRADE*, because, by construction, the true sample mean is close to 0. This happens because portfolio re-balancing increases the weight of a stock (when an asset manager buys more of that stock), but this comes at the expense of using that same weight to decrease the holdings of other stocks in the same portfolio.

net buys by 0.7-3.0 percent for a 1-point increase in ratings and this effect corresponds to 2.7%-11.5% of one standard deviation in the change in portfolio weights.

To ensure that our results are robust, in Table 3, we repeat all the previous analyses using quadratic and cubic RDD specifications. The comprehensive set of results tabulated under these alternative higher-order RDD models is quantitatively and qualitatively similar to the results obtained under our baseline linear RDD specification.

To put our RDD results into perspective, we replace the RDD analysis with an OLS estimation of the association between online consumer satisfaction scores and fund trades. Specifically, we use the initial (pre-RDD) fund-level sample of 1,327,439 fund-stock-year-quarter observations to perform an OLS test instead of exploiting an RDD test based on our final rounding-based fund-level sample of 592,889 fund-stock-year-quarter observations. In the online appendix Table OA2, we document that unrounded online consumer satisfaction ratings are positively associated with mutual fund trades. Based on the coefficient size range of 0.027-0.038 (p-value<0.05), we interpret this effect as an increase in mutual fund trades by 12.6%-17.8% of one standard deviation in the change in portfolio weights (i.e., 0.027/0.214 [lower bound] to 0.038/0.214 [upper bound]) for a one-unit increase in online consumer satisfaction scores (*POSITIVE (unrounded*)).[40] Given that *POSITIVE (unrounded*)/100 has a 0-1 range, a one-unit increase in *POSITIVE(unrounded)* corresponds, on a 0-100 percentage-point scale, to an increase in customer ratings by 100 percentage-points. Therefore, a 100-percentage-point increase in online consumer satisfaction scores (on a 1-100 score range) is associated with an increase in mutual fund trades by 12.6%-17.8% of one standard deviation in the change in portfolio weights. A 1-percentage-point increase in these scores (on a 1-100 score range) would therefore translate

[40] The standard deviation of *FUND TRADE* in the larger, pre-RDD fund-level sample is 0.214 (the equivalent of this in the RDD-based fund-level sample was 0.261).

to an increase in mutual fund trades by 0.13%-0.18% of one standard deviation in the change in portfolio weights.

Consistent with the econometric view that correlated omitted variable bias can either inflate or deflate the magnitude of the effect, or even make effects go in the opposite direction than expected (see, e.g., Angrist and Pischke, 2009, section 3.2.2; Wilms, Mäthner, Winnen and Lanwehr, 2021), the economic magnitude under OLS in the pre-RDD sample is much smaller than that of the baseline RDD estimate.[41] Recall that *POSTIVE(unrounded)* is an endogenous measure of consumer tweets; so, the OLS coefficient on this variable is affected by the abovementioned correlated omitted variable bias problem. The above issue notwithstanding, our OLS results overall qualitatively align with our RDD results; hence, all the evidence presented across both OLS and RDD tests suggests that mutual fund managers, on average, increase their allocation of weights to stocks in their portfolios when such stocks receive favorable consumer satisfaction ratings.

4.2 Cross-sectional Variation: The Effect of Information Asymmetry

In this subsection, we investigate the link between consumer tweets and mutual fund trades by examining scenarios or mechanisms where the positive association between consumer tweets and the allocation of weights to portfolio stocks varies in size.

We argue that in the presence of information asymmetries, social media-based information such as consumer purchase tweets should be useful to asset managers in alleviating their uncertainties about portfolio firms' prospects. We, therefore, expect the positive association between consumer tweets and fund trades to be more pronounced when information asymmetry is high. To this test this, we adopt three proxies of information asymmetry: (i) the extent of

[41] Recall that, in the RDD sample, we had documented that a 1-percentage-point exogenous increase in these online ratings (on a 1-100 scale) induces an increase in fund trade by 2.7%-11.5% of one standard deviation in the change in portfolio weights.

disagreement between analysts' earnings forecasts and firms' actual earnings (proxied by *FORECAST_ERROR*, which is measured as the absolute value of the difference between analysts' consensus earnings [EBIT] forecasts and actual firm earnings for a given firm-quarter, deflated by total assets); (ii) the one-step-ahead volatility of earnings (proxied by *CF_UNCERTAINTY*, which is measured as the one-quarter-ahead standard deviation of earnings [EBITDA], deflated by average assets from the previous 4 quarters); and (iii) the extent of accounting quality (proxied by *ABS_ACCRUALS*, which we measure as the principal component factor of the absolute value of residuals from 4 accruals models estimated in the quarterly setting [i.e., Jones model (Jones, 1991), modified Jones model (Dechow et al., 1995), DD model (Dechow and Dichev, 2002) and Collins et al.'s (2017) non-linear sales growth model].[42] By construction, these proxies are increasing in the degree of information asymmetry.

The results reported in Table 4 show that the coefficient on the interaction term is positive and significant across all tabulations, but the coefficient on the standalone term $ABOVE\_VS\_BELOW_{i,q}$ is insignificant. The evidence is in alignment with our expectation that our baseline finding should be more pronounced when the information asymmetry is more severe. We therefore conclude that social media-based information helps to alleviate information asymmetries between managers of portfolio firms and mutual fund managers.

4.3 Robustness Checks

Two sources of heterogeneity in asset managers' responses could potentially arise at *individual* cutoff points (0.5, 1.5, 2.5, …, 99.5) and/or the *pooled* cutoff point of 0 (i.e., the pooling together of all individual cutoff points into a single average cutoff). First, one may worry that how asset managers respond to changes in online consumer satisfaction ratings may differ across *individual*

---

[42] Collins et al. (2017) show the quarterly approach to estimating several of these accruals models.

rounding thresholds. For example, an asset manager is likely to perceive a 1-point increase from 49 to 50 (where cutoff=49.5) differently than a 1-point increase from 89 to 90 (where cutoff=89.5) and so on. That is, some individual cutoffs tend to have more psychological weight than others when it comes to asset managers' portfolio re-balancing decision-making. Second, the visual evidence we earlier presented under RDD plots (based on maximum and larger bandwidths) suggests that the slope of the RDD plot *below* the pooled cutoff point differs from that *above* the pooled cutoff point. This implies that potential heterogeneity exists in asset managers' responses to changes in online consumer satisfaction ratings below vs. above the *pooled* cutoff point. We therefore control for these two sources of heterogeneities in asset managers' responses via, respectively, augmenting our baseline RDD model with individual rounding thresholds fixed effects (see, e.g., Derrien et al. 2023), as well as controlling for the slope of the RDD plot (see, e.g., Hollenbeck et al. 2019). As in the RDD literature, we proxy for the slope of the RDD with *ABOVE_VS_BELOW* × *POSITIVE (unrounded)*. The results, reported in Table 5, Columns 1-4, yield RDD estimates (i.e., coefficients on *ABOVE_VS_BELOW* in the alternative setups) that are qualitatively similar to our baseline estimates, implying that our results are robust.[43]

As additional robustness check, we use an alternative RDD specification that specifies side by side as the main right-hand side variables, *ABOVE_VS_BELOW* and the running variable—*RATING_RELATIVE_TO_THRESHOLD* (the running variable is used instead of the earlier-specified *POSITIVE (unrounded)* in our baseline RDD model). For this alternative setup, we additionally control for the above-mentioned sources of heterogeneities in asset managers' responses; this time round, the slope is proxied with *ABOVE_VS_BELOW* ×

[43] As in existing RDD studies, we neither interpret the coefficient on the endogenous variable POSITIVE (unrounded), nor the coefficient on the slope term, because researchers more often than not tend to have no expectations about these statistics (see, e.g., the above-cited studies). The focus of an RDD model is therefore solely to isolate and fixate on causal effects, which is captured alone by the RDD indicator.

RATING_RELATIVE_TO_THRESHOLD. In these alternative setups tabulated as Columns 5-8 of Table 5, we continue to find a significant coefficient on the standalone term *ABOVE_VS_BELOW*. Our results are therefore robust to different ways of specifying an RDD.

Finally, to lend credibility to the feature that the application of exogenous rounding thresholds in the format XX.5 tends to displace online ratings, which asset managers then respond to, we perform falsification tests based on multiple placebo rounding thresholds. The idea is to show that, asset managers do not respond to changes in online ratings that were rounded using placebo (fake) thresholds. To this end, we adopt the placebo thresholds that take the form of XX.5±0.1, XX.5±0.15 and XX.5±0.2 (i.e., thresholds in the form XX.3, XX.35, XX.4, XX.6, XX.65, and XX.7 are alternatively used as placebo cutoff points).

We then use these new cutoff points to round up vs. round down exact (unrounded) ratings and redefine *ABOVE_VS_BELOW* for each of these placebo threshold formats. We do not adopt placebo thresholds that take the form XX.5±0.3, XX.5±0.35, and XX.5±0.4, because they yield significantly very little to no variation in the reconstructed (placebo) RDD indicator—*ABOVE_VS_BELOW* for these placebos (recall that, the use of the real threshold XX.5 yielded a 51% variation in the RDD indicator [see Table 1]).[44]

In Table 6, we re-estimate our baseline model replacing *ABOVE_VS_BELOW* with a placebo-based RDD indicator under alternative setups that adopt different placebo thresholds. Interestingly, in none of these tabulations do we find that using placebo thresholds to displace online ratings leads to marked changes in asset managers' decision-making. Specifically, the coefficient of *ABOVE_VS_BELOW* is statistically insignificant under all of the placebo cases

[44] Econometrically, exploiting very-little to no variations in the RDD indicator would create a situation where the RDD indicator is subsumed in our baseline model Eq. (1) (we confirm this in untabulated tests).

adopted. This observation thus suggests that the economic effects that we document are a result of the application of *actual* (and not fake) rounding rules.

4.4 Actively-Managed Equity Mutual Funds vs. Passive Funds

In this subsection, we distinguish between two categories of funds: *active* vs. *passive*. Following Kacperczyk, Sialm and Zheng (2008), we identify the universe of *actively*-managed U.S. open-end equity mutual funds covered by the standard Center for Research in Security Prices (CRSP) survivorship-bias-free Mutual Fund Database. We observe 1,802 unique actively-managed equity mutual funds in our all-funds sample of 4,051 funds used in our baseline tests. From the slice that is left after isolating active equity funds, we can specifically identify index funds and so use this as a representative sample of *passive* funds.

In general, managers of active funds engage in extensive research, analysis, and management activities seeking to outperform the market, so there is high portfolio turnover resulting in rapid portfolio re-balancing. In contrast, managers of passive funds play a limited role in the above asset management activities, as there is minimal portfolio turnover—passive funds tend to only track the performance of the underlying stocks in a pre-specified index. Consequently, one would expect frequent portfolio re-balancing in response to changes in consumer tweets in active funds, but not in passive funds. It is for this reason that we earlier regarded the category of passive funds as a placebo in our setup.

To test the above conjecture, we re-estimate our baseline results here using the sub-samples of actively-managed equity funds vs. passive funds. Similar to our baseline estimations, we employ a regression analysis under the RDD setup to estimate the causal effects of consumer tweets on fund trades and quantify the exact economic impact of consumer satisfaction ratings for these actively-managed funds vs. index funds. As before, our multivariate estimation approach allows us to account for any possible factors that may introduce bias, including attributes of portfolio firms, as well as variations over time and across various funds.

In Table 7, Panel A, we present the results of our linear RDD estimation for the active vs. passive funds categories. As per the results tabulated, we find the coefficient on the RDD indicator to be positive (coeff. ranges between 0.010 and 0.011) and significant at the 10% level for the active funds category (we however do not observe statistically significant results for the passive funds category). As before, we interpret the coefficient of 0.010-0.011 as an increase in net buys by 1.0-1.1 percent in response to an exogenous, rounding-induced 1-point increase (on a scale of 0-100) in Twitter-based consumer satisfaction scores. Overall, the results obtained for actively-managed equity funds systematically mimic those earlier observed for our all-funds-sample, but we observe no such results for the placebo group of passive funds. This therefore provides comfort that our tests identify, consistent with practitioner insights, the category of funds that are actually making use of alternative data—active funds.

To further demonstrate robustness of our results obtained for actively-managed equity funds vs. passive funds, we repeat all of the analyses above, this time, using quadratic and cubic RDD specifications in Table 7, Panel B. The comprehensive set of results tabulated under these alternative setups is overall quantitatively and qualitatively similar to the results obtained under our baseline linear RDD specification.

5. Information and Mutual Fund Returns

We now investigate whether the information content of Twitter-based consumer satisfaction scores has a causal impact on the performance of mutual funds. We start by constructing a measure of asset managers' reliance on LikeFolio data. We then use this measure to test the extent to which asset managers' reliance on online customer ratings affects mutual fund performance. Consistent with the literature, all of our performance tests are based on the restricted sample of actively-managed equity funds.

5.1 Construction of Asset Managers' Rate of Adoption of LikeFolio Data

We now describe the methodology adopted to construct our measure of reliance on consumer tweets (hereafter, "*RELIANCE_TWITTER*"). In the same spirit as the Kacperczyk and Seru (2007) reliance on public information ("RPI") measure, we construct *RELIANCE_TWITTER* based on the sensitivity of fund managers' portfolio holdings to LikeFolio's Twitter (now X)-based consumer satisfaction scores.

We estimate *RELIANCE_TWITTER* using a two-step procedure. In the first step, we find how much of the average percentage change in a fund's quarterly holdings of a stock in the current period—i.e., $\Delta FUND\_TRADE_{f,i,q}$—can be attributed to the quarterly change in *POSITIVE(unrounded)* for that same stock in the last period—i.e., $\Delta POSITIVE(unrounded)_{i,q}$. Alternatively, as *ABOVE_VS_BELOW* essentially represents a 1-percentage-point increase in customer satisfaction/happiness scores, we also alternatively estimate how much of $\Delta FUND\_TRADE$ can be attributed to an exogenous, 1-percentage-point increase in customer satisfaction/happiness scores (captured by *ABOVE_VS_BELOW*). Specifically, for each fund and period, we estimate the cross-sectional regression using all stocks in the fund's portfolio similar to Kacperczyk and Seru (2007), but this time replace Kacperczyk and Seru's (2007) right-hand side variable—change in analyst's consensus recommendation—with $\Delta POSITIVE(unrounded)$, or alternatively *ABOVE_VS_BELOW*.

In the second step, we set *RELIANCE_TWITTER* equal to the unadjusted R^2 of the regression from step 1. [45] In sum, *RELIANCE_TWITTER*, which captures the extent of asset

[45] Refer to Kacperczyk and Seru (2007) for a detailed discussion on the construction of their so-called reliance on public information (RPI) measure.

managers' reliance on LikeFolio Twitter-based data, equals the unadjusted R^2 of the regression of Δ*FUND_TRADE* on Δ*POSITIVE (unrounded*) for each fund. We are, therefore, interested only in the degree by which fund managers rely on LikeFolio data in their portfolio decisions. Thus, only R^2 from the above regression matters for constructing our measure.

We start with a confirmatory check. We perform a cross-sectional pooled regression using the entire sample of fund-periods to get a sense of whether $\Delta FUND\_TRADE_{f,i,q}$ is truly positively associated with (and therefore sensitive) to $\Delta POSITIVE(unrounded)_{i,q}$ in the entire sample. We find this to be the case in Table 8, Panel A, where the coefficient on $\Delta POSITIVE(unrounded)_{i,q}$ /100 is 0.062 (p-value<0.01). Satisfying ourselves with the existence of a positive correlation, we then proceed to estimate the same regression, this time, for each fund-year-quarter, allowing us to construct the *RELIANCE_TWITTER* measure for each fund-year-quarter. Our sample exhibits significant cross-sectional variation in *RELIANCE_TWITTER*. The (untabulated) mean value of *RELIANCE_TWITTER* equals 23.3% with a standard deviation of 32.3% and a range between 0.01% and 99.9%. To examine the anatomy of our measure, we examine the relationship between *RELIANCE_TWITTER* and fund characteristics such as Size, Age, Expense, and Turnover. To achieve this objective, we sort funds into quintiles based on the *RELIANCE_TWITTER* level for each period. Then, we calculate the average values of the chosen variables for each portfolio. This process is repeated for all subsequent time periods. Finally, we compute the time-series average by aggregating all the cross-sectional averages.

The resulting values of these fund characteristic variables from the above-described sorting procedure are presented in Table 8, Panel B. They clearly reveal a monotonic relationship between *RELIANCE_TWITTER* (a measure of the extent of reliance on LikeFolio data) and funds' total net assets. Specifically, smaller funds exhibit a greater dependence on social media information compared to larger funds. This relationship can be explained by two plausible explanations. Firstly, it is possible that larger funds, which typically enjoy a stronger reputation and offer higher wages, employ more skilled managers. These managers may thus rely less on third-party-provided information (such as LikeFolio data) and instead utilize alternative sources such as proprietary research. This could explain the lower reliance on social media information among larger funds (Kacperczyk and Seru, 2007). Alternatively, the lower *RELIANCE_TWITTER* observed in larger funds may be attributed to their challenges in trading based on online information. Due to their substantial size, executing trades may have a significant price impact, making it difficult for them to take advantage of online information effectively (Berk and Green, 2004). This suggests that the utilization of online information may come with increased costs and trading activity. Furthermore, we find a negative relationship between the age of funds and *RELIANCE_TWITTER*. This implies that younger funds tend to rely more heavily on online information compared to their more established counterparts. The final row of Table 8, Panel B provides a summary of the pairwise correlations between *RELIANCE_TWITTER* and other fund characteristics, measured using standard coefficients of correlation. All the statistics presented are highly statistically significant.

5.2 Mutual Fund Performance and Reliance on Consumer Tweets

We argued that the utilization of LikeFolio data, which represents a novel and emerging source of information, should provide fund managers with an informational advantage. To empirically examine this hypothesis, we employ a panel regression model, estimating the following regression equation:

$$\alpha_{f,q} = \beta_0 + \beta_1 RELIANCE\_TWITTER_{f,q-1} + \gamma Controls_{f,q-1} + \varepsilon_{f,q} \quad , \qquad (2)$$

where $\alpha_{f,q}$ denotes the performance measure of fund f at time q. Mutual fund performance is measured using the four-factor alpha of Carhart (1997), estimated as the difference between the realized fund return in excess of the risk-free rate and the expected excess fund return from a four-factor model, which includes the market, size, value, and momentum factors. In estimating the factor loadings, we use rolling-window time-series regressions of fund return over the prior three years[46]. $RELIANCE\_TWITTER_{f,q}$ denotes reliance on social media (Twitter) for fund *f* at time q. In the regression specification, we also control for fund characteristics related to fund performance that might affect the RELIANCE_TWITTER-performance relationship.

The multivariate regression framework concurrently controls for these different characteristics. The vector of $Controls$ includes: (i) Kacperczyk and Seru (2007) reliance on public information (RPI) measure, which equals the unadjusted R^2 of the regression of percentage changes in fund managers' portfolio holdings (Δ*FUND_TRADE*) on changes in analysts' past recommendations; (ii) fund size, measured as the natural logarithm of fund total net assets under management; (iii) fund age, measured as the natural logarithm of 1 plus fund's age in years; (iv) expenses, measured as the expense ratio of a fund; (v) turnover, measured as the turnover ratio of a fund; (vi) flows, measured as the growth rate of assets under management after adjusting for the appreciation of the fund's assets; and (vii) team, which is an indictor variable if a fund is team-managed. γ is therefore a vector of coefficients that correspond to these variables. We include time, fund family, fund style and fund manager FEs and we cluster by fund and time[47].

We expect a positive and significant β_1. The results of this regression are reported in Table 9, with the Carhart four-factor risk adjusted alpha measure in Panel A, the Fama and MacBeth (1973) regressions in Panel B and regressions involving an alternative measure of *RELIANCE_TWITTER* in Panel C. The results across all five columns of Panel A of Table 9 show

[46] We also measure fund performance using Fama-French three-factor risk/style adjusted alpha (Fama and French, 1993). We present this result in the internet appendix.
[47] Following Petersen (2009), we consider two-way clustering by both funds and time.

that, as expected, *RELIANCE_TWITTER* is positively related to fund performance (as proxied by Carhart four-factor alpha). The results are statistically significant and robust to the inclusion of fund characteristics. They are also economically significant. The coefficients range from 0.019-0.038 with t-stats ranging from 2.29-3.86 depending on the model specification. For instance, in Column 1, a one-standard deviation increase in *RELIANCE_TWITTER* increases the risk-adjusted alpha of Carhart (1997) by 1.21% (0.038*0.32) per year. To rule out the possibility of Kacperczyk and Seru (2007) RPI effect, we control for RPI in our regressions. We find our results to be both statistically and economically significant and not subsumed by the RPI effect. This suggests that our measure captures fundamentally something different from what is contained in analysts' consensus recommendations.

Next, we perform a Fama and MacBeth (1973) cross-sectional estimation. Panel B of Table 9 presents the coefficient estimates from the cross-sectional regression of four-factor risk adjusted alpha on *RELIANCE_TWITTER* while controlling for the same variables used in the panel regression. We find that *RELIANCE_TWITTER* is positively related to fund performance (as captured by the Carhart four-factor risk adjusted alpha) and we find this result to be statistically significant at 1% level.

In addition to our primary measure of reliance on LikeFolio-supplied Twitter-based customer satisfaction scores, we introduce an alternative measure using the unadjusted R^2 of the regression of Δ*FUND_TRADE* on the indicator variable for whether a firm's customer satisfaction score is just above or below rounding thresholds (Recall that, the RDD indicator represents a 1-point increase in consumer satisfaction scores on a scale of 1-100; so, the RDD indicator is essentially also a change variable). We label this variable *RELIANCE_TWITTER (II)*. A key strength of this analysis relative to those immediately above is that it allows us to exploit exogeneity in constructing our measure of reliance on LikeFolio data, thereby alleviating concerns about correlated omitted variables. The findings related to this alternative measure are presented in Panel C of Table 9. Our analysis reveals that when utilizing this alternative measure, the results remain consistent with those obtained from our primary measure. Specifically, we observe both statistical significance and economic magnitude similar to those reported in Panel A of Table 9. These consistent results strengthen our conclusion that mutual funds with a greater reliance on alternative data sources exhibit the potential for earning abnormally high future returns.

5.3 Holding-based Performance

We now provide a robustness check based on holding-based measures of fund performance. Holding-based measures offer a valuable approach for addressing potential model misspecification issues in factor regressions. To assess fund performance, we utilize the holding-based measures introduced by Daniel, Grinblatt, Titman, and Wermers (1997). These measures decompose a fund's overall excess return into three distinct components: "Characteristic Selectivity" (CS), "Characteristic Timing" (CT), and "Average Style" (AS). The CS measure captures a fund's ability to select securities with favorable characteristics, independent of the fund's overall investment style. The CS represents a measure of the stock picking ability of the fund manager. The CT measure evaluates a fund's timing decisions related to the buying and selling securities based on their characteristics. The CT measure captures the style-timing ability of the fund manager by examining any additional performance generated by exploiting time-varying expected returns of the size, book-to-market, or momentum benchmark portfolios. The AS measure reflects the average performance of a fund's investment style, considering both the CS and CT components.[48]

[48] The computation of these measures is described in more detail in Daniel, Grinblatt, Titman and Wermers (1997).

Employing these holding-based measures allows us to separately evaluate a fund's selectivity, timing, and average style, providing a nuanced understanding of the factors contributing to its overall performance. By examining the relationships between *RELIANCE_TWITTER* and these holding-based measures, we contribute to the understanding of how reliance on Twitter-based customer satisfaction scores impacts different dimensions of fund performance.

The panel regression results are reported in Table 10. The dependent variables are the characteristic-based benchmark adjusted measures, whereas our independent variable remains *RELIANCE_TWITTER*. We maintain the same control variables, fixed effects, and clustering as used in Table 9. Results for CS, CT, and AS measures are reported in Panels A, B, and C, respectively. In Panel A, we find *RELIANCE_TWITTER* to be positively and significantly related to CS across Column 1-4 except for Column 5. These results on CS largely suggest that funds with a greater reliance on Twitter-based customer satisfaction scores have superior performance that can be attributed to fund managers' stock-picking abilities. In Panel B of Table 10, we also find *RELIANCE_TWITTER* to be positively related to CT, these results though significant across Column 1-3, but are weaker compared to the CS results. This notwithstanding, the CT results themselves seem to offer some support that the superior performance by funds with a higher reliance on Twitter-based customer satisfaction scores can be partly explained by fund managers' timing abilities. Lastly, Panel C of Table 10 reports results using AS. We find that *RELIANCE_TWITTER* is not significantly related to AS across different model specifications except for Column 1. This points to weak evidence that our performance effects are explained by AS. Overall, putting all of these results together, our evidence seems to suggest that funds with a greater reliance on social media signals outperform the market because fund managers have better stock-picking and market-timing abilities.

6. Lifespan of the Value of Information

How lasting is the above-determined value of LikeFolio data to asset managers? Through their stock-picking and timing abilities, is there persistence of the above-documented gains to asset managers? To address this focal question, we examine the temporal dynamics of the relationship between fund managers' reliance on customer satisfaction scores and the resulting performance gains. Specifically, we investigate how quickly mutual fund managers can react to the information contained in these scores and how long the benefits derived from this alternative data can be sustained. By delving into the speed and longevity of the gains, we can have a deeper understanding of the practical implications and potential limitations of incorporating Twitter-based customer satisfaction scores into investment strategies. Such insights would provide valuable guidance to fund managers seeking to leverage this new, alternative data source effectively.

To investigate the speed and duration of gains derived by fund managers through their reliance on Twitter-based customer satisfaction scores, we conduct a dynamic regression and present the findings in Table 11. Our analysis involves regressing the Carhart four-factor alpha on the interaction between RELIANCE_TWITTER and time dummies while controlling for relevant variables and fixed effects, similar to the approach used in Table 9, Panel A. Our key variables of interest are the interactions between *RELIANCE_TWITTER* and year-quarter dummies. Using 2012Q1 as our regression benchmark, we find a significant positive relationship between *RELIANCE_TWITTER* and fund performance only for the period from 2012Q4 to 2013Q2. Our analysis thus reveals that fund managers who rely on Twitter-based customer satisfaction scores extracted gains from 2012Q4 to 2013Q2 at most. However, starting from 2013Q3 and extending to 2015Q4, we find no significant gains associated with funds' reliance on customer satisfaction scores.

This period aligns with LikeFolio's initiation of major publicity of their database in mainstream media in 2013Q3 (see Figure 3). As more fund managers and the general public (e.g., retail investors) begin to notice and utilize the data (see Figure 4),[49] the previously observed gains quickly saturate, and we even observe instances (e.g., 2015Q4) where fund managers relying on customer ratings begin to experience losses. These findings highlight the time-sensitive nature of the gains from relying on Twitter-based customer satisfaction scores. The results suggest that the trading gains associated with this alternative data source are not sustainable over an extended period. Moreover, the increased adoption of the database by fund managers appears to erode the initial advantages, ultimately leading to diminishing returns and potential losses.

In Table 12, we use a market-based performance measure (i.e., 90-days-ahead cumulative equally-weighted abnormal returns at each firm-year-quarter-end) instead of the one-quarter-ahead performance reported (realized) by mutual funds and find evidence that is overall qualitatively similar: we observe performance effects in 2012Q4 of the pre-publicity period and not in the publicity or post-publicity periods.

As we find evidence that asset managers make money only in the pre-publicity period, implying that only early-adopters of the LikeFolio data (i.e., asset managers who privately accessed the data before it became public) profited from their trades, we next attempt to provide specific evidence on the performance-determining trades by the early-adopters of the LikeFolio data. Specifically, we decompose our measure of trade—*FUND_TRADE*—into *FUND_BUY* vs.

---

[49] Figure 4 shows the CRSP trading volume increases steadily from the pre-publicity period (before 2013Q3) to the commencement of the publicity period (2013Q3), but then increases a bit more sharply, spiking towards the end of the publicity period (2014Q1) before falling.

FUND_SELL, which are the two components of *FUND_TRADE*. By construction, *FUND_BUY (FUND_SELL)* is increasing (decreasing) in *FUND_TRADE*; so, *FUND_BUY* has a positive value, whereas *FUND_SELL* has a negative value. Consequently, the higher the absolute value of *FUND_BUY (FUND_SELL*), the higher the propensity of a portfolio manager to buy (sell) a stock in a portfolio.

We then identify early-reliance funds in two steps. As a first step, we fixate on pre-publicity *RELIANCE_TWITTER* values and determine the median of *RELIANCE_TWITTER* during the pre-publicity period (we focus on the pre-publicity period, because this is the period where asset managers make money and hence the period where one can identify early adopters of the LikeFolio data). In the second step, we flag early adopters as those with *RELIANCE_TWITTER* values that are above the above-determined median during the pre-publicity period. We then average fund buy vs. sell per year-quarter and graph the time series of these averages in Figure 5.

The visual evidence presented in Figure 5 shows a dramatic spike in *FUND_SELL* in 2013Q1 relative to 2012Q4 (compared to the movement in *FUND_BUY* for the same periods). Given that our mutual fund performance measures are forward-looking (measured as a one-quarter-ahead proxy) and given further that we had earlier observed (based on both measures of future performance) that asset managers initially made money in the quarter following 2012Q4 (i.e., the one-quarter-ahead performance loads in 2012Q4),[50] we conclude that the same period in which we observed the dramatic spike in *FUND_SELL* (relative to *FUND_BUY*)—i.e., 2013Q1—is the same period that we had observed (based on both measures of performance) that asset managers initially made money from their trades.

[50] Performance is one-quarter ahead; so, if this forward-looking measure significantly loads in 2012Q4, this implies that asset managers initially made money not in 2012Q4, but rather the quarter after that—i.e., 2013Q1.

Overall, we observe performance only before publicity, implying that only early adopters of the LikeFolio data profited from their trades. These traders essentially privately leveraged the LikeFolio data and made money before the public noticed the data and began leveraging the data in their trades. Our analysis underscores the importance of considering the temporal dynamics and competitive landscape when evaluating the value of alternative data sources in generating long-term investment gains.

7. Conclusion

Unlike other products, information loses its value when more people use it (Admati and Pfleiderer, 1995, 1996, 1997). This study examines the speed at which the information content of alternative data is impounded into the trading process and the duration of its value. We focus on a new dataset offered on the market ("LikeFolio") and investigate whether mutual fund managers use such information, the speed at which the information content of this data is impounded into the trading process, and the duration of the value of such data to its users.

Using a regression discontinuity design (RDD), we document that mutual fund managers rely on customer-generated comments about a company's products and services on social media platforms in making portfolio allocation decisions. Mutual funds rely more on social media signals when information asymmetry is more pronounced as proxied by greater analysts' forecast errors, more volatile cash flows, and poorer accounting quality. Funds that rely more on social media signals yield higher abnormal future returns and exhibit better stock-picking and market-timing abilities.

Lastly, we find that funds with a greater reliance on social media signals earn higher abnormal returns only when they have private access to such data; the abnormal returns dissipate when the data becomes public.

References

Admati, A., 1985. A Noisy Rational Expectations Equilibrium for Multi-Asset Securities Markets, Econometrica, 629-657.

Admati, A., Pfleiderer, P., 1986. A Monopolistic Market for Information, Journal of Economic Theory, 400-438.

Admati, A., Pfleiderer, P., 1987. Viable Allocations of Information in Financial Markets, Journal of Economic Theory, 76-115.

Admati, A., Pfleiderer, P., 1988. Selling and Trading Information in Financial Markets, American Economic Review, Paper and Proceedings, 96-103.

Admati, A., Pfleiderer, P., 1990. Direct and Indirect Sale of Information, Econometrica, 901-928.

Aiken, A.L., Clifford, C.P., Ellis, J.A., 2015. Hedge funds and discretionary liquidity restrictions, Journal of Financial Economics 116, 197-218.

Ali, A., Chen, X., Yao, T., Yu, T., 2008. Do mutual funds profit from the accruals anomaly?. Journal of Accounting Research, 46(1), 1-26.

Altonji, J.G., Elder, T.E., Taber, C.R., 2005. An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling. Journal of Human Resources 40 (4): 791–821.

Angrist, J.D., Pischke, J.S., 2009. Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Baker, M., Litov, L., Wachter, J.A., Wurgler, J., 2010. Can mutual fund managers pick stocks? Evidence from their trades prior to earnings announcements. Journal of Financial and Quantitative Analysis, 45(5), 1111-1131.

Barber, B.M., Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, Review of Financial Studies 21, 785-818.

Bartov, E., Faurel, L., Mohanram, P., 2018. Can Twitter help predict firm-level earnings and stock returns? The Accounting Review, 93, 25-57.

Berk, J., Green, R., 2004. Mutual Fund Flows and Performance in Rational Markets, Journal of Political Economy, 112, 6, 1269-1295.

Best, M.C., Cloyne, J.S., Ilzetzki, E., Kleven, H.J., 2020. Estimating the elasticity of intertemporal substitution using mortgage notches. The Review of Economic Studies, 87(2), 656-690.

Blankespoor, E., Miller, G.S., White, H.D., 2014. The role of dissemination in market liquidity: Evidence from firms' use of Twittertm. The Accounting Review, 89, 79–112.

Bollen, J., Mao, H., Zheng, X., 2011. Twitter mood predicts the stock market. Journal of Computational Science, 2, 1–8.

Brown, N.C., Wei, K.D,. Wermers, R., 2014. Analyst recommendations, mutual fund herding, and overreaction in stock prices. Management Science, 60(1), 1-20.

Bushee, B., M., Jung, and G. Miller. 2011. Conference presentations and the disclosure milieu. Journal of Accounting Research 49 (5): 1163292.

Bushee, B.J., Jung, M.J,. Miller, G.S., 2017. Do investors benefit from selective access to management?. Journal of Financial Reporting, 2(1), pp.31-61.

Calonico, S., Cattaneo, M.D., Titiunik, R., 2015. Optimal data-driven regression discontinuity plots. Journal of the American Statistical Association 110: 1753-1769.

Chen, H., De, P., Hu, Y., Hwang, B., 2014. Wisdom of crowds: the value of stock opinions transmitted through social media. Rev. Financ. Stud. 27, 1367–1403 .

Chuprinin, O., Gaspar, S., Massa, M., 2019. Adjusting to the information environment: News tangibility and mutual fund performance, Management Science 65, 1430-1453.

Cohen, L., Frazzini, A., Malloy, C., 2008. The small world of investing: Board connections and mutual fund returns. Journal of Political Economy 116 (5): 951279.

Collins, D.W., Pungaliya, R.S., Vijh, A.M., 2017. The effects of firm growth and model specification choices on tests of earnings management in quarterly settings. The Accounting Review, 92(2), 69-100.

Crowley, R., Huang, W., Lu, H., 2018. Discretionary disclosure on Twitter. Working Paper.
Curtis, A., Richardson, V., Schmardebeck, R., 2016. Investor Attention and the Pricing of Earnings News. Handbook of Sentiment Analysis in Finance (Chapter 8): 212–232.
Daniel, K., Grinblatt, M., Titman, S., Wermers, R., 1997. Measuring Mutual Fund Performance with Characteristic-Based Benchmarks. Journal of Finance, 52, 1035-58.
Dechow, P.M., Dichev, I.D., 2002. The quality of accruals and earnings: The role of accrual estimation errors. The accounting review, 77(s-1), 35-59.
Dechow, P.M., Sloan, R.G., Sweeney, A.P., 1995. Detecting earnings management. Accounting review, 193-225.
Derrien, F., Garel, A., Petit-Romec, A., Weisskopf, J.P., 2023. Online reputation and debt capacity. Journal of Financial and Quantitative Analysis, 1-41
Domo, 2021. Data never sleeps. Available at https://www.domo.com/solution/data-never-sleeps-6 (accessed Sep. 15 2023).
Dugast, J., Foucault, T., 2021. Equilibrium data mining and data abundance. HEC Paris Research Paper No. FIN-2020-1393, Université Paris-Dauphine Research Paper, (3710495).
Engelberg, J.E., Parsons, C.A., 2011. The causal impact of media in financial markets, Journal of Finance 66, 67-97.
Fama, E.F., French, K.R., 2008. Dissecting anomalies. The Journal of Finance 63(4), pp.1653-1678.
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests, Journal of Political Economy 81, 607-636.
Fama, E.F., French, K., 1993. Common risk factors in the returns on bonds and stocks, Journal of Financial Economics 33, 3–56.
Fang, L., Kosowski, R., 2007. Comparing Stars – Trading on Star Mutual Funds' Holdings and Star Analysts' Recommendations. Working paper, Tanaka Business School, Imperial College London.
Fang, L.H., Peress, J., Zheng, L., 2014. Does media coverage of stocks affect mutual funds' trading and performance?, Review of Financial Studies 27, 3441-3466.
Ferson, W.E., Khang, K., 2002. Conditional performance measurement using portfolio weights: Evidence for pension funds, Journal of Financial Economics 65, 249–282.
Franck, A., Kerl, A., 2013. Analyst forecasts and European mutual fund trading. Journal of Banking & Finance, 37(8), pp.2677-2692.
GASM, 2016. GSAM Perspectives: The Role of Big Data in Investing. Available at https://www.gsam.com/content/gsam/global/en/market-insights/gsam-insights/gsam-perspectives/2016/big-data/gsam-roundtable.html (accessed Aug. 9th, 2023).
Gleason, C. A., Lee, C. M., 2003. Analyst forecast revisions and market price discovery. The Accounting Review, 78(1), 193-225.
Goldstein, I., Spatt, C.S., Ye, M., 2021. Big data in finance. The Review of Financial Studies, 34(7), 3213-3225.
Gu, Z., Li, Z., Yang, Y.G., Li, G., 2019. Friends in need are friends indeed: An analysis of social ties between financial analysts and mutual fund managers. The Accounting Review, 94(1), 153-181.
Hollenbeck, B., Moorthy, S., Proserpio, D., 2019. Advertising strategy in the presence of reviews: An empirical analysis. Marketing Science, 38, 793-811.
Huang, J., 2018. The customer knows best: The investment value of consumer opinions, Journal of Financial Economics 128, 164-182.
Huang, R., 2020. The Financial Consequences of Customer Satisfaction: Evidence from Yelp Ratings and SBA Loans. Available at SSRN 3064343.
Jame, R., 2018. Liquidity provision and the cross section of hedge fund returns, Management Science 64, 3288-3312.
Jones, J.J., 1991. Earnings management during import relief investigations. Journal of accounting research, 29(2), 193-228.
Jung, M.J., Naughton, J.P., Tahoun, A., Wang, C., 2018. Do firms strategically disseminate? Evidence from corporate use of social media. The Accounting Review, 93, 225-252.

Kaniel, R., Parham, R., 2017. WSJ category kings - the impact of media attention on consumer and mutual fund investment decisions, Journal of Financial Economics 123, 337-356.
Kacperczyk, M., Seru, A., 2007. Fund manager use of public information: New evidence on managerial skills. Journal of Finance 62, 485-528.
Kelley, E., Tetlock, P., 2013. How wise are crowds? insights from retail orders and stock returns. J. Finance 68, 1229–1265 .
Mao, Y., Wei, W., Wang, B., Liu, B., 2012. Correlating S&P 500 stocks with Twitter data. Proceedings of the 1st ACM International Workshop on Hot Topics on Interdisciplinary Social Networks: 69–72.
Martin, I.W., Nagel, S., 2022. Market efficiency in the age of big data. Journal of Financial Economics, 145(1), pp.154-177.
McLean, R.D., Pontiff, J., 2016. Does academic research destroy stock return predictability? Journal of Finance 71, 5–32.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703
Lee, L.F., Hutton, A.P., Shu, S., 2015. The role of social media in the capital market: Evidence from consumer product recalls. Journal of Accounting Research, 53, 367–404.
Lee, C., Ma, P., Wang, C., 2015. Search-based peer firms: aggregating in- vestor perceptions through internet co-searches. J. Financ. Econ. 116, 410–431 .
Liberti, J.M., Petersen, M.A., 2019. Information: Hard and soft. Review of Corporate Finance Studies, 8, 1–41.
Peress, J., 2014. The media and the diffusion of information in financial markets: Evidence from newspaper strikes. the Journal of Finance, 69(5), 2007-2043.
Petersen, M.A., 2009. Estimating standard errors in finance panel data sets: Comparing approaches, Review of Financial Studies 22, 435-480.
Petusevsky, S., 2010. The Whole Foods Market Cookbook: A Guide to Natural Foods with 350 Recipes. New York: Clarkson Potter.
Parida, S., Teo, T., 2016. The impact of more frequent portfolio disclosure on mutual fund performance. Journal of Banking and Finance 87: 427–65.
Pinnuck, M., 2003. An examination of the performance of the trades and stock holdings of fund managers: Further evidence. Journal of Financial and Quantitative Analysis 38(4), 811-828.
Roberts, M.R., Whited, T.M., 2013. Endogeneity in empirical corporate finance1. In Handbook of the Economics of Finance (Vol. 2, pp. 493-572). Elsevier.
Solomon, D., Soltes, E., 2015. What are we meeting for? The consequences of private meetings with investors. Journal of Law and Economics 58 (2): 325-55.
Solomon, D., Soltes, E., Sosyura, D., 2014. Winners in the spotlight: Media coverage of fund holdings as a driver of flows, Journal of Financial Economics 113, 53-72.
Sprenger, T.O., Tumasjan, A., Sandner, P.G., Welpe, I.M., 2013. Tweets and trades: The information content of stock microblogs. European Financial Management, 20, 926–957.
Tang, V.W., 2018. Wisdom of crowds: Cross-sectional variation in the informativeness of third-party-generated product information on Twitter. Journal of Accounting Research, 56, 989-1034.
Teo, M., 2011. The liquidity risk of liquid hedge funds, Journal of Financial Economics 100, 24-44.
Tetlock, P.C., 2007. Giving content to investor sentiment the role of media in the stock market, Journal of Finance, 1139-1168.
Walker, D., 1990. Customer First: A Strategy for Quality Service. Aldershot: Gower.
Webster Jr., F.E., 1994. Executing the new marketing concept. Marketing Management, Vol. 3 No. 1 pp. 9-18.
Wilms, R., Mäthner, E., Winnen, L., Lanwehr, R., 2021. Omitted variable bias: a threat to estimating causal relationships. Methods in Psychology, 5, 100075.

Appendix A

Variable definition

Variable	**Description**
FUND_TRADE	This measures the change in the weight (w) of stock i from quarter $q-1$ to quarter q in fund manager f*'s* portfolio (Pinnuck 2003, p.814). $FUND\_TRADE_{f,i,q} = w_{f,i,q} - w^{pq}_{f,i,q-1}$, where $w_{f,i,q} = \frac{P_{i,q} H_{f,i,q}}{\sum_{i=1}^{N}(P_{i,q} H_{f,i,q})}$; $w^{pq}_{f,i,q-1} = \frac{P_{i,q} H_{f,i,q-1}}{\sum_{i=1}^{N}(P_{i,q} H_{f,i,q-1})}$; $P_{i,q}$=price of stock i at quarter q; $H_{f,i,q}$=split-adjusted number of shares of stock i held by fund f at quarter q; and N= the number of different stocks held by each fund manager.
<u>LikeFolio Product Info</u>	
POSITIVE (unrounded)	LikeFolio's "0-100 scale" customer satisfaction/happiness score, measured as one hundred (100) times "the ratio of the number of tweets that convey a positive assessment of products and brands over the number of tweets that convey a non-neutral (either positive or negative) assessment of products and brands" (Tang 2018, p.991)
ABOVE_VS_BELOW	A dummy equal to 1 (0) if a given firm-year-quarter trade observation is just above (below) customer satisfaction/ happiness score rounding cutoff points (thresholds), with assignment into groups derived by identifying which LikeFolio "0-100 scale" unrounded, customer satisfaction/ happiness scores are close to any of the following rule-of-thumb rounding thresholds: 0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, …, 97.5, 98.5, and 99.5.
<u>Firm Characteristics</u>	
SIZE_ASSETS	The natural logarithm of total assets (ATQ) in millions of US dollars.
TANGIBILITY_NET	Property, Plant and Equipment (PPENTQ)/ total assets (ATQ). Net PPE [PPENTQ] rather than Gross PPE [PPEGTQ] is used because quarterly data has several missing observations for PPEGTQ.
LEVERAGE	(Debt in current liabilities [DLCQ] + long-term debt [DLTTQ])/ assets (ATQ)
PROFITABILITY	Earnings before interest and tax (EBITQ)/assets (ATQ)
OPERATING_CYCLE	log ((receivables [RECTQ] / sales [SALEQ] + inventory [INVTQ] / cost of sales [COGSQ]) *360)
MARKET_TO_BOOK	(market value of equity [PRCCQ × CSHOQ] + book value of debt [ATQ - CEQQ]) / total assets [ATQ] {see Lin, Ma, Malatesta, and Xuan, 2013; Smith 2016}
Z_SCORE	A modified version of Altman's (1968) Z-score calculated as: (3.3 × (earnings before interest and tax [EBITQ] / total assets [ATQ]) + 1 × (Sales [SALEQ]/ total assets [ATQ]) + 1.2 × (Current Assets [ACTQ]/ total assets [ATQ]) + 1.4 × (Retained Earnings [REQ]/ total assets [ATQ]) {see Smith 2016}
LOSS	A dummy variable that assumes the value of 1 if firm reports accounting loss; 0 otherwise
DIVIDEND	A dummy equal to 1 if the firm paid a dividend (i.e., if DVQ>0); 0 otherwise. Note that DVQ is calculated as the difference in current year-to-date dividend and last quarter year-to-date dividend [i.e., ΔDVY]
SALES_GROWTH	[SALEQ q - SALEQ q-1] / SALEQ q-1
INVENTORY	Inventories (INVTQ) / assets (ATQ)
CASH_HOLDINGS	Cash and cash equivalents (CHEQ) / assets (ATQ)

Fund Characteristics	
RELIANCE_TWITTER	This measure captures the extent of reliance on social media (Twitter) and equals the unadjusted R^2 of the regression of percentage changes in fund managers' portfolio holdings on changes in POSITIVE (unrounded) customer satisfaction/happiness score, constructed in the spirit of Kacperczyk and Seru 2007 RPI measure
RELIANCE_TWITTER (II)	This measure captures the extent of reliance on social media (Twitter) and equals the unadjusted R^2 of the regression of percentage changes in fund managers' portfolio holdings on ABOVE_VS_BELOW, {a dummy equal to 1 (0) if a given firm-year-quarter trade observation is just above (below) customer satisfaction/ happiness score rounding cutoff points (thresholds)}, constructed in the spirit of Kacperczyk and Seru (2007) RPI measure
RPI	This measures the extent of reliance on public information (RPI) and equals the unadjusted R^2 of the regression of percentage changes in fund managers' portfolio holdings on changes in analysts' past recommendations (Kacperczyk and Seru 2007)
Fund Size (TNA)	This is the total net asset of fund in millions of dollars
Fund Age	This is the natural log of one plus the fund age in years which represents the number of years a fund is present in the CRSP equity mutual fund database
Expenses	Represents a fund's expense ratio
Turnover	Is the turnover ratio of a fund
Flow	This is the growth rate of assets under management after adjusting for the appreciation of the fund's assets (Sirri and Tufano 1998)
Team	Is an indicator variable that equals 1 if a fund is managed by multiple managers and 0 otherwise
Performance measures:	
(fund-level performance)	
4 factor alpha $_{f,q+1}$	The 4-factor alpha of Carhart (1997), estimated as the difference between the *realized* fund return in excess of the risk-free rate and the expected excess fund return from a 4-factor model, which includes the market, size, value, and momentum factors. In estimating the factor loadings, we use rolling-window time-series regressions of fund return over the prior three years.
(stock-level performance)	
CAR $_{i,q+1}$	The 90-days-ahead cumulative equally-weighted abnormal returns of firm *i*'s stock at year-quarter-end of period *q*, where abnormal returns at each stock-date is calculated based on Fama and French's (2008) size- and book-to-market-based 5*5 sorting methodology.
DGTW Holding-based measures	
Characteristic Selectivity (CS)	Represents a measure of stock selection ability of the fund manager constructed according to Daniel, Grinblatt, Titman and Wermers (1997)
Characteristic Timing (CT)	Captures the style-timing ability of the fund manager constructed according to Daniel, Grinblatt, Titman and Wermers (1997)
Average Style (AS)	Represents a measure of average style selection skill of the fund manager constructed according to Daniel, Grinblatt, Titman and Wermers (1997)

Figure 1A. LikeFolio app (free access)

This figure shows the LikeFolio app (free access) that displays rounded consumer satisfaction ratings. This rounded rating scores are freely accessible to the public (including asset managers), who can observe real-time moving changes in buyers' consumer satisfaction/happiness. In these figures, handwritten-numbers 1, 2, 3 correspond to the steps that an app user typically takes to input a ticker ID and to search for a company associated with that ticker ID.

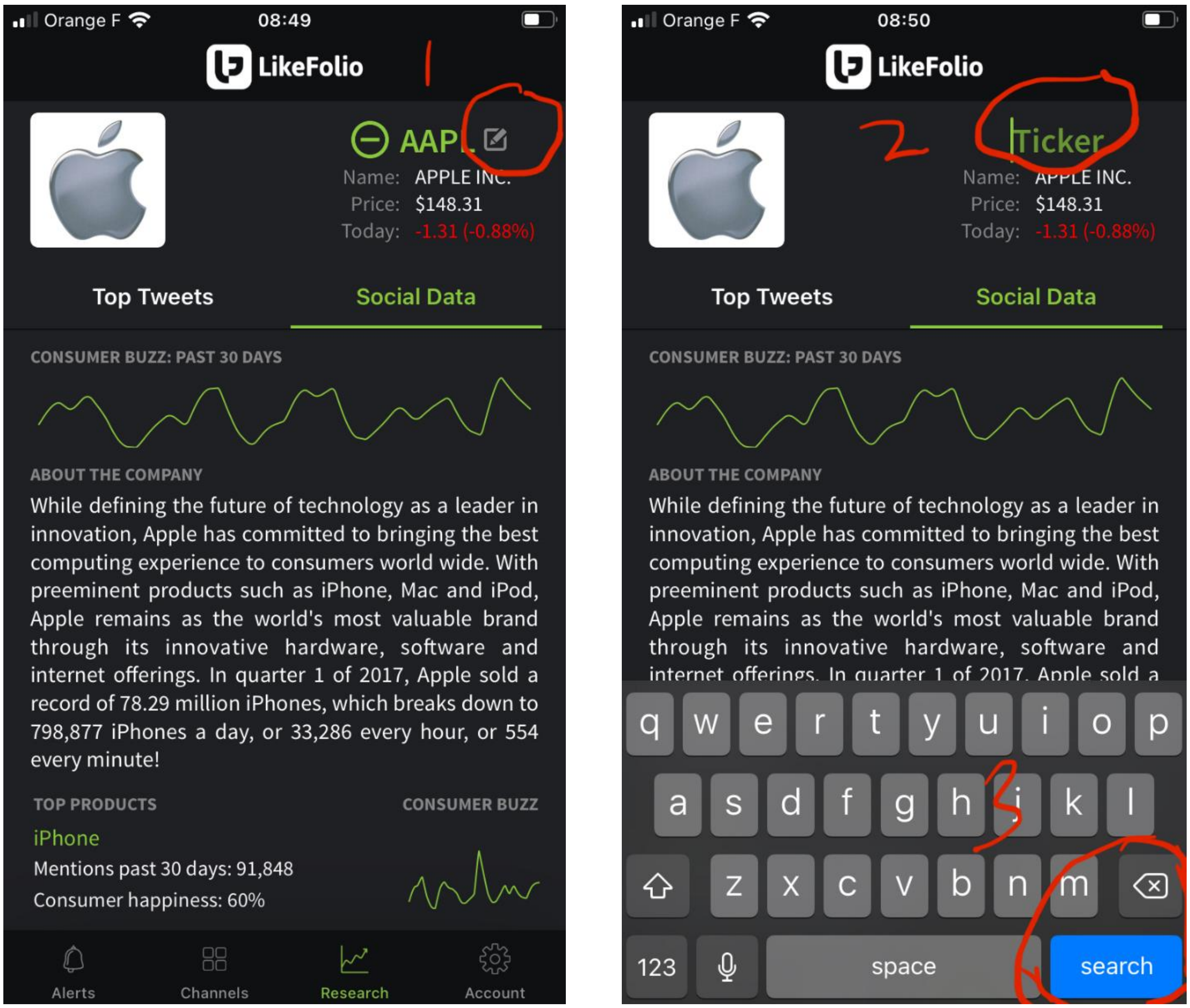


Figure 1B. **LikeFolio dashboard (paid subscription)**

This figure shows the LikeFolio dashboard (paid subscription) that displays rounded consumer satisfaction ratings.

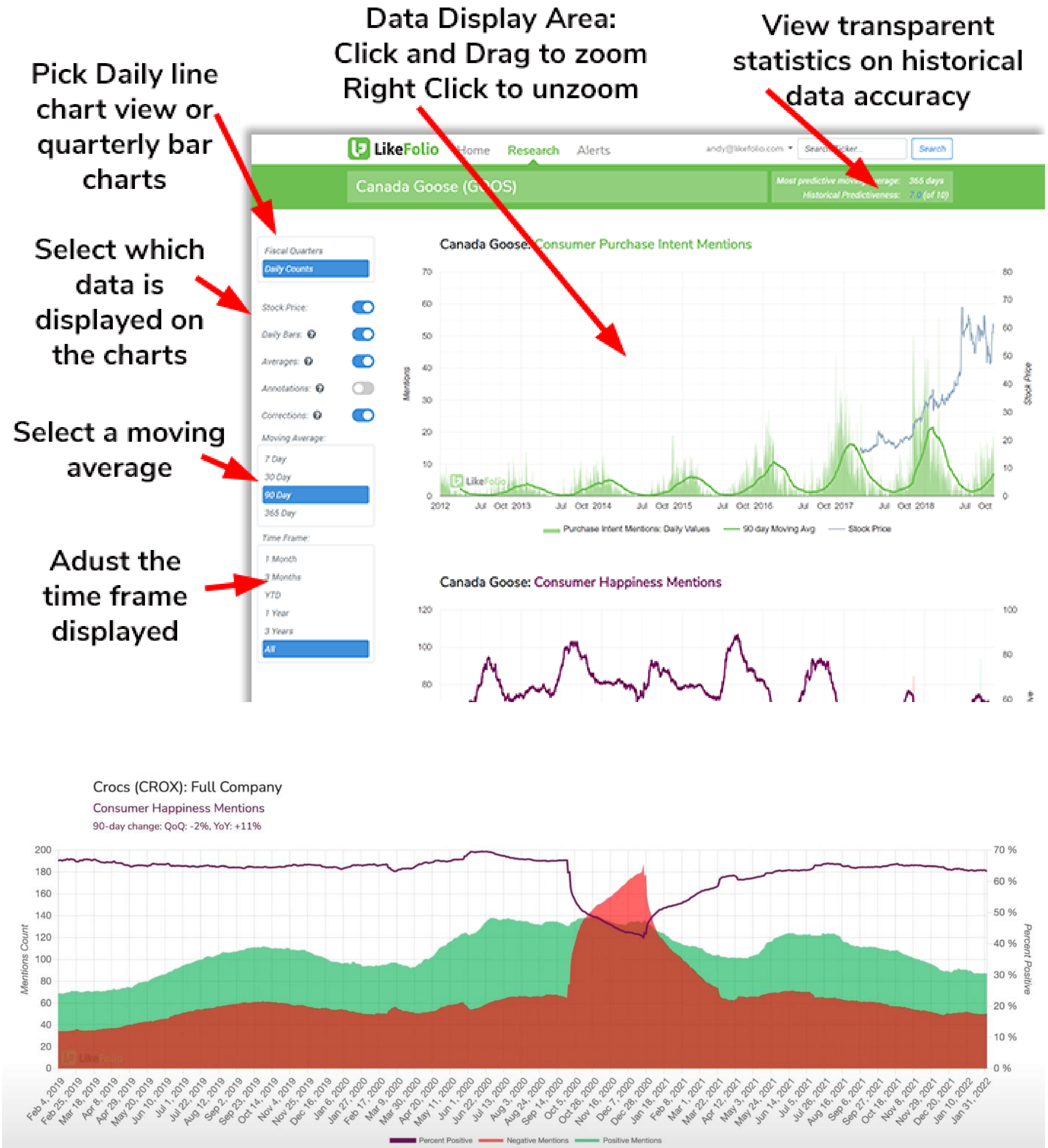
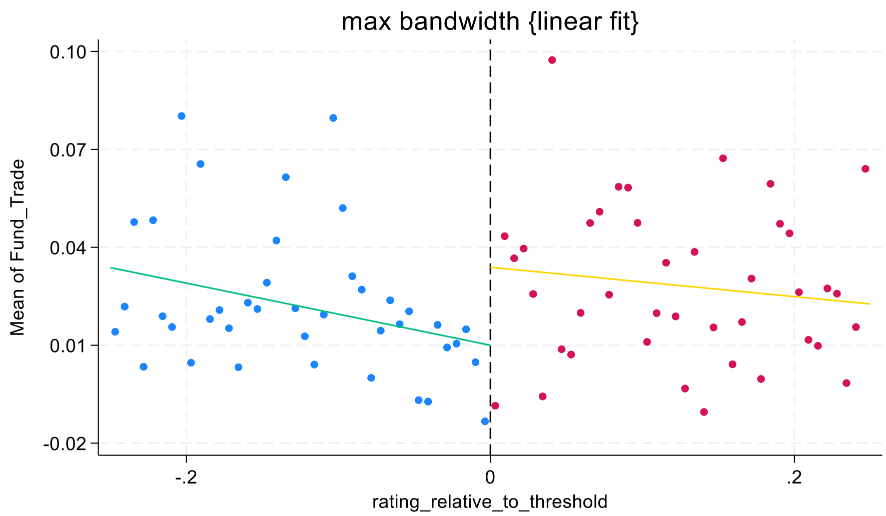
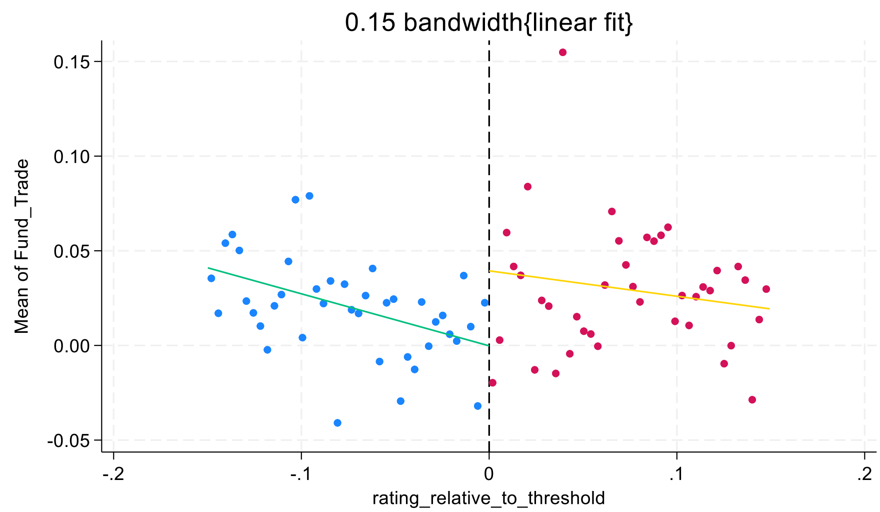
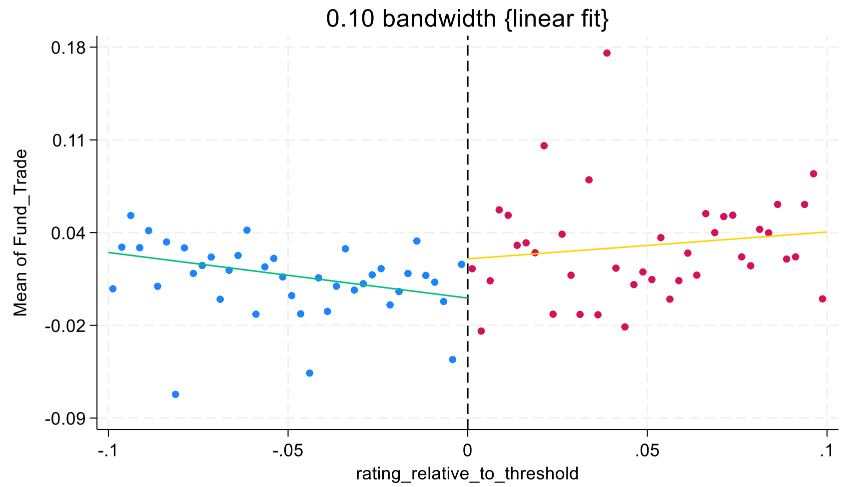


Figure 2: RDD plot

In these figures, we plot average fund trade as a function of LikeFolio customer satisfaction ratings around cutoff points. We follow the classic approach of dividing the customer satisfaction ratings space into an even selected vs. optimal number of bins on the left and right sides of cutoff points. We start with the bins at the pooled cutoff point of 0 (i.e., the position of any of the individual cutoff points), and gradually work our way out to the right and left of 0 in order to ensure that no bin contains both observations just above (i.e., treatment group) and observations just below (i.e., control group) cutoff points. For each bin, we compute the average fund trade and the corresponding midpoint value of POSITIVE and plot the average fund trade for each bin on the vertical axis against the corresponding midpoint POSITIVE value at each bin on the horizontal axis. We re-center fund trade observations around their respective cutoff points to 0 to allow for easily identifying the position of each fund trade observation relative to any of the 100 rounding thresholds.






0.10 bandwidth {linear fit}
Mean of Fund_Trade
0.18
0.11
0.04
-0.02
-0.09
-.1
-.05
0
.05
.1
rating_relative_to_threshold

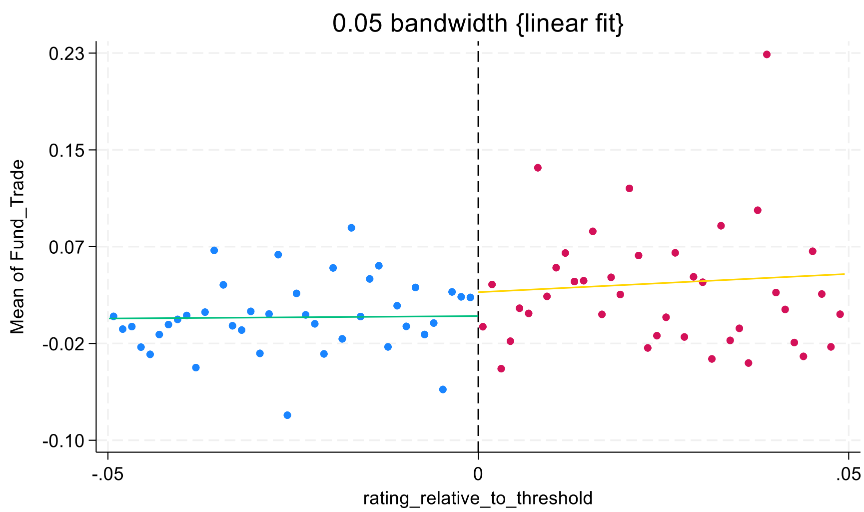
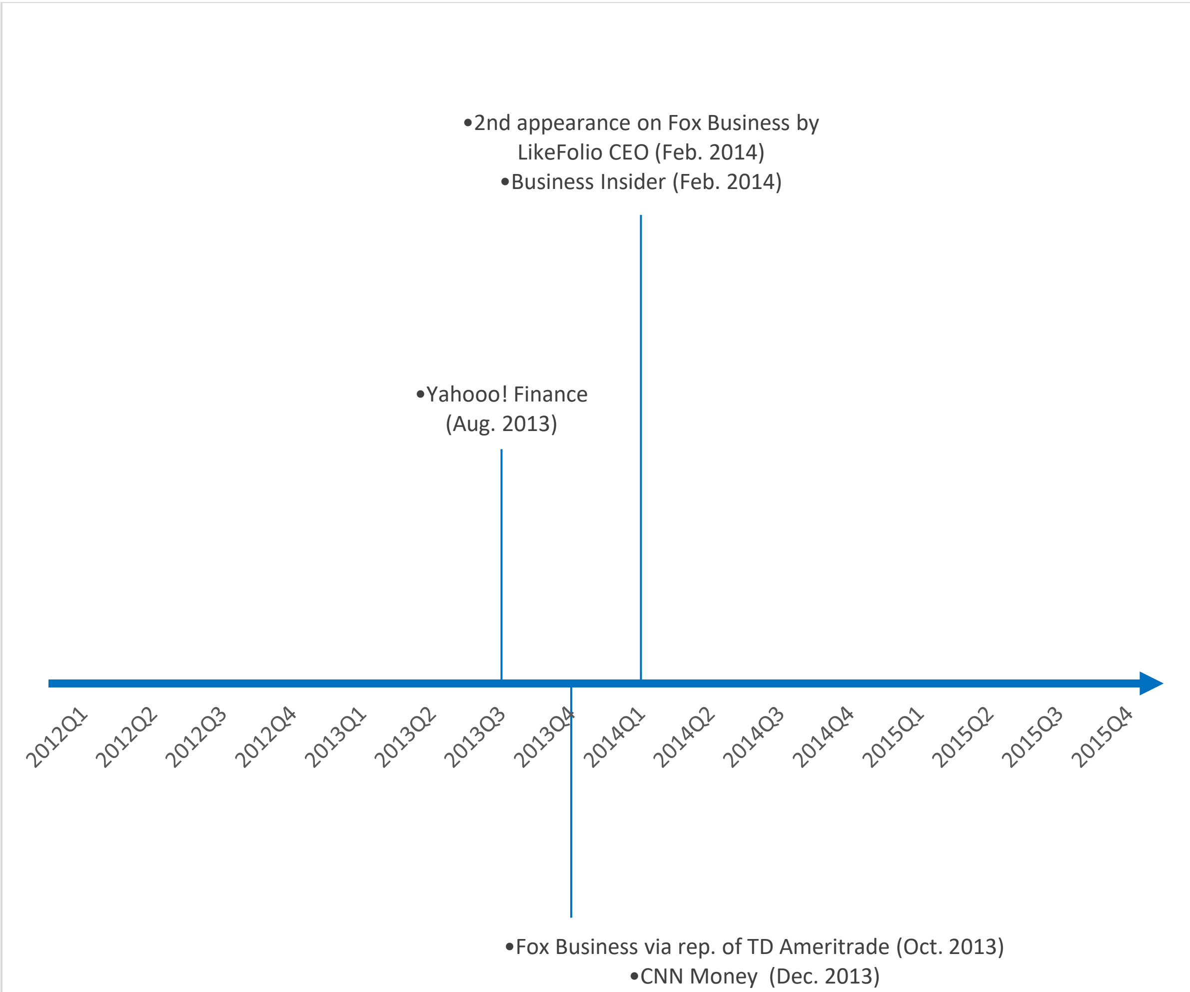

0.05 bandwidth {linear fit}
Mean of Fund_Trade
0.23
0.15
0.07
-0.02
-0.10
-.05
0
.05
rating_relative_to_threshold

Figure 3. Timeline of LikeFolio publicity events



News sources (last accessed 03 Jan. 2024):

- *Yahoo! Finance:*
  https://www.yahoo.com/now/2013-08-14-invest-in-what-you-know-goes-social-via-new-web-si.html
- *Fox Business (1st appearance):*
  LikeFolio Facebook video post: https://www.facebook.com/LikeFolio/videos/562544580468123/
- *CNN Money:*
  https://money.cnn.com/gallery/pf/2013/12/11/best-financial-apps/2.html
- *Fox Business (2nd appearance):*
  LikeFolio YouTube video post: https://www.youtube.com/watch?v=Sy3ArhAMf-U
- *Business Insider:*
  LikeFolio Facebook video post: https://www.facebook.com/LikeFolio/videos/627813323941248

Figure 4: Time series of trading volume

This figure presents a graphical depiction of the trend of quarterly CRSP trading volume of LikeFolio-covered firms in our sample.

Figure 5: Time series of the average fund buy vs. sell for early adopters with a greater reliance on LikeFolio data

This figure presents a graphical depiction of the trend of quarterly FUND_BUY vs. FUND_SELL—the components of FUND_TRADE. The graph is intended to analyze the trading behavior of funds that likely profit from early-reliance on LikeFolio data before the data becomes public. To this end, we identify early-reliance funds in two steps. As a first step, we fixate on pre-publicity RELIANCE_TWITTER values and determine the median of RELIANCE_TWITTER during the pre-publicity period (we focus on the pre-publicity period, because this is region where one is able to identify early adopters of the LikeFolio data). We then, in a second step, flag early-adopters as those with RELIANCE_TWITTER values that are above the above-determined median during the pre-publicity period. We then average FUND_BUY vs. FUND_SELL per year-quarter and graph the time series of these averages.

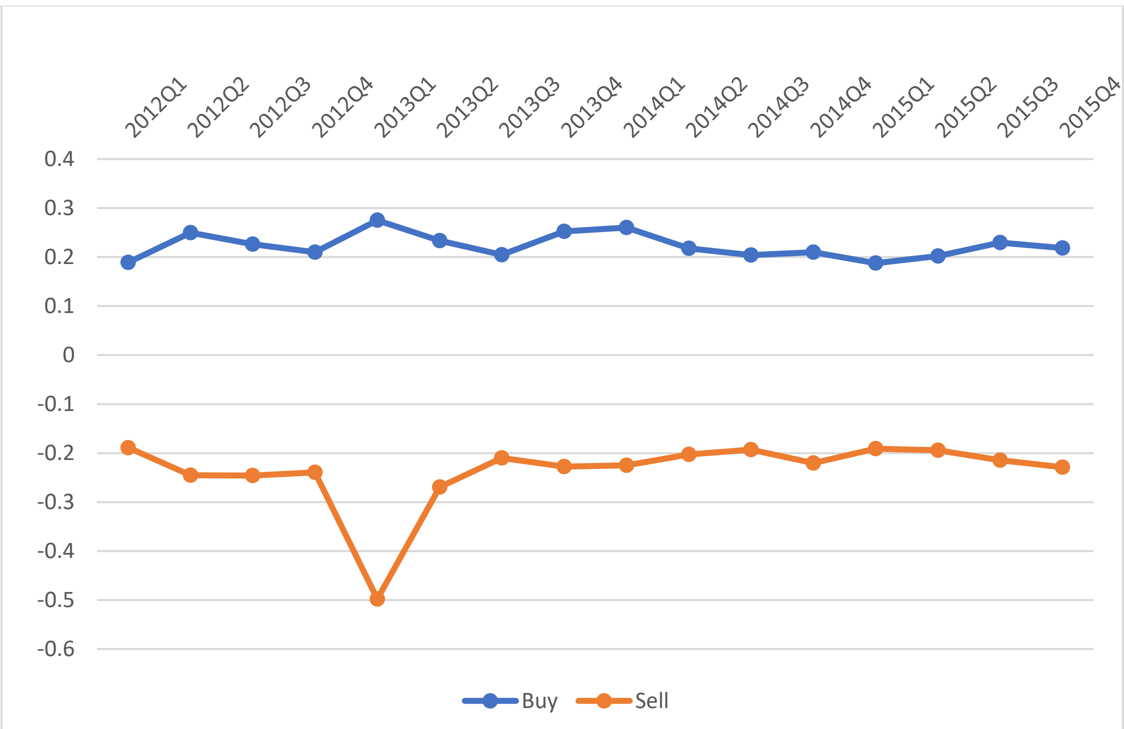


Table 1: Descriptive statistics

Panel A: Summary statistics of fund trade and LikeFolio data

This table reports the descriptive statistics of fund trade for all funds in the CRSP mutual fund database and also for the active equity funds having their holdings covered by LikeFolio. The table also reports the descriptive statistics of original customer sentiment (POSITIVE{unrounded}/100), rounding thresholds, and ABOVE_VS_BELOW [See Appendix A for detailed definitions of variables].

	ALL FUND STATS						
	Mean	Std Dev	5th Pctl	25th Pctl	Median	75th Pctl	95th Pctl
No. of Funds	4051						
FUND_TRADE	0.025	0.261	-0.417	-0.032	0.000	0.036	0.428
	ACTIVELY-MANAGED EQUITY FUND STATS						
	Mean	Std Dev	5th Pctl	25th Pctl	Median	75th Pctl	95th Pctl
No. of Funds	1802						
FUND_TRADE	0.034	0.302	-0.419	-0.080	0.000	0.094	0.568
	FIRM STATS (#obs=2,509)						
	Mean	Std Dev	5th Pctl	25th Pctl	Median	75th Pctl	95th Pctl
POSITIVE (unrounded)/100	0.85	0.12	0.64	0.79	0.87	0.95	0.99
Rounding Thresholds	85.1	12.3	63.5	78.5	87.5	94.5	99.5
ABOVE_VS_BELOW	0.51	0.50	0.00	0.00	1.00	1.00	1.00

Panel B: Covariate balance in firm characteristics across RDD groups

This table reports mean values of all firm-level characteristics in the group ABOVE_VS_BELOW=1 versus the group ABOVE_VS_BELOW=0 and performs a statistical test of the difference in these means [See Appendix A for detailed definitions of variables]. Asterisk demarcations *, **, and *** represent statistical significance at 10%, 5%, and 1% levels, respectively.

Variable	ABOVE_VS_BELOW=1			ABOVE_VS_BELOW=0			Difference in Means
		(1)	(2)		(3)	(4)	(3) - (1)
	N	Mean	Median	N	Mean	Median	t-val.
SIZE_ASSETS	1,272	8.158	8.172	1,237	8.137	8.304	-0.27
PROFITABILITY	1,272	0.014	0.023	1,237	0.018	0.023	1.24
TANGIBILITY_NET	1,272	0.235	0.166	1,237	0.232	0.162	-0.39
LEVERAGE	1,272	0.271	0.239	1,237	0.269	0.229	-0.17
MARKET_TO_BOOK	1,272	2.541	1.894	1,237	2.531	1.960	-0.12
Z_SCORE	1,272	0.746	1.004	1,237	0.735	1.042	-0.15
SALES_GROWTH	1,272	0.061	0.023	1,237	0.047	0.016	-1.03
INVENTORY	1,272	0.099	0.048	1,237	0.098	0.043	-0.32
LOSS	1,272	0.227	0.000	1,237	0.213	0.000	-0.83
OPERATING_CYCLE	1,272	5.708	5.807	1,237	5.760	5.846	1.34
CASH_HOLDINGS	1,272	0.188	0.117	1,237	0.185	0.111	-0.32
DIVIDEND	1,272	0.491	0.000	1,237	0.517	1.000	1.30

Table 2: Linear RDD model

This table reports estimates of a linear regression of FUND_TRADE on ABOVE_VS_BELOW - the indicator for fund trade observations just above vs. below rounding cutoff points [See Appendix A for detailed definitions of variables]. Asterisk demarcations *, **, and *** represent statistical significance at 10%, 5%, and 1% levels, respectively. The table reports heteroskedasticity-adjusted t-values in parentheses, observations are clustered at the stock level, and the regression controls for fund and year-quarter FEs.

	Panel A: Fund trade (max bandwidth)			
	(1)	(2)	(3)	(4)
cons	0.004	-0.007	-0.138***	-0.108***
	(0.28)	(-0.50)	(-4.15)	(-3.11)
POSITIVE (unrounded) /100	0.021	0.034*	0.056***	0.050***
	(1.24)	(1.94)	(3.13)	(2.82)
ABOVE_VS_BELOW	0.007*	0.008*	0.007*	0.008*
	(1.70)	(1.85)	(1.79)	(1.77)
SIZE_ASSETS			0.012***	0.010***
			(5.06)	(3.83)
PROFITABILITY			0.122	0.126
			(1.32)	(1.53)
TANGIBILITY_NET			-0.005	-0.009
			(-0.40)	(-0.89)
LEVERAGE			-0.014	-0.006
			(-1.60)	(-0.81)
MARKET_TO_BOOK			0.007***	0.006***
			(4.50)	(4.12)
Z_SCORE			-0.001	-0.002
			(-0.61)	(-0.82)
SALES_GROWTH			-0.013	-0.012
			(-1.34)	(-1.34)
INVENTORY			0.015	0.025
			(0.87)	(1.62)
LOSS			0.004	0.001
			(0.50)	(0.19)
OPERATING_CYCLE			-0.002	-0.004
			(-0.86)	(-1.39)
CASH_HOLDINGS			0.048*	0.040
			(1.67)	(1.44)
DIVIDEND			-0.015**	-0.010*
			(-2.50)	(-1.66)
Year-Qtr FE	NO	YES	NO	YES
Fund FE	NO	YES	NO	YES
Cluster by Stock	YES	YES	YES	YES
N	593030	592889	593030	592889
Adjusted R^2	0.000	0.062	0.009	0.066

Table 2 (Cont'd)

	Panel B: Fund trade (Bandwidth =< 0.15)			
	(1)	(2)	(3)	(4)
cons	-0.002	-0.025	-0.155***	-0.136***
	(-0.13)	(-1.31)	(-3.77)	(-3.01)
POSITIVE (unrounded) /100	0.026	0.053**	0.065***	0.070***
	(1.17)	(2.32)	(2.71)	(2.90)
ABOVE_VS_BELOW	0.010*	0.011**	0.010*	0.010**
	(1.85)	(2.00)	(1.91)	(2.00)
Control Variables	NO	NO	YES	YES
Year-Qtr FE	NO	YES	NO	YES
Fund FE	NO	YES	NO	YES
Cluster by Stock	YES	YES	YES	YES
N	347300	347083	347300	347083
Adjusted R^2	0.001	0.072	0.008	0.075

	Panel C: Fund trade (Bandwidth =< 0.10)			
	(1)	(2)	(3)	(4)
cons	-0.002	-0.019	-0.158***	-0.126**
	(-0.10)	(-0.81)	(-3.01)	(-2.25)
POSITIVE (unrounded) /100	0.019	0.039	0.059**	0.057**
	(0.73)	(1.44)	(2.08)	(2.02)
ABOVE_VS_BELOW	0.022***	0.021***	0.020***	0.019***
	(3.09)	(3.17)	(3.09)	(3.24)
Control Variables	NO	NO	YES	YES
Year-Qtr FE	NO	YES	NO	YES
Fund FE	NO	YES	NO	YES
Cluster by Stock	YES	YES	YES	YES
N	248146	247871	248146	247871
Adjusted R^2	0.002	0.079	0.010	0.082

	Panel D: Fund trade (Bandwidth =< 0.05)			
	(1)	(2)	(3)	(4)
cons	-0.047	-0.050*	-0.209**	-0.176**
	(-1.43)	(-1.82)	(-2.44)	(-2.20)
POSITIVE (unrounded) /100	0.062	0.067**	0.082**	0.064*
	(1.60)	(2.02)	(2.01)	(1.88)
ABOVE_VS_BELOW	0.030**	0.026**	0.023**	0.022***
	(2.33)	(2.54)	(2.50)	(2.65)
Control Variables	NO	NO	YES	YES
Year-Qtr FE	NO	YES	NO	YES
Fund FE	NO	YES	NO	YES
Cluster by Stock	YES	YES	YES	YES
N	120067	119750	120067	119750
Adjusted R^2	0.004	0.095	0.013	0.100

Table 3: Higher order RDD models

This table reports estimates of a quadratic and cubic regressions of FUND_TRADE on ABOVE_VS_BELOW - the indicator for fund trade observations just above vs. below rounding cutoff points [See Appendix A for detailed definitions of variables]. Asterisk demarcations *, **, and *** represent statistical significance at 10%, 5%, and 1% levels, respectively. The table reports heteroskedasticity-adjusted t-values in parentheses, observations are clustered at the stock level, and the regression controls for fund and year-quarter FEs. For brevity, estimations here are performed using the maximum bandwidth.

	Quadratic Model				Cubic Model			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
cons	-0.050	-0.041	-0.116***	-0.107**	0.068	0.074	-0.057	-0.022
	(-1.26)	(-1.08)	(-2.62)	(-2.29)	(1.00)	(1.03)	(-0.79)	(-0.30)
POSITIVE (unrounded) /100	0.164	0.123	-0.008	0.047	-0.450	-0.472	-0.314	-0.389
	(1.53)	(1.24)	(-0.09)	(0.52)	(-1.19)	(-1.28)	(-0.81)	(-1.03)
square {POSITIVE (unrounded) /100}	-0.092	-0.058	0.041	0.002	0.869	0.873	0.522	0.686
	(-1.31)	(-0.89)	(0.71)	(0.04)	(1.37)	(1.47)	(0.86)	(1.16)
cubic {POSITIVE (unrounded) /100}					-0.470	-0.455	-0.236	-0.335
					(-1.45)	(-1.52)	(-0.79)	(-1.15)
ABOVE_VS_BELOW	0.007*	0.008*	0.007*	0.008*	0.007*	0.008*	0.007*	0.008*
	(1.73)	(1.85)	(1.78)	(1.76)	(1.74)	(1.87)	(1.78)	(1.78)
Control Variables	NO	NO	YES	YES	NO	NO	YES	YES
Year-Qtr FE	NO	YES	NO	YES	NO	YES	NO	YES
Fund FE	NO	YES	NO	YES	NO	YES	NO	YES
Cluster by Stock	YES	YES	YES	YES	YES	YES	YES	YES
N	593030	592889	593030	592889	593030	592889	593030	592889
Adjusted R^2	0.000	0.062	0.009	0.066	0.000	0.062	0.009	0.066

Table 4: The moderating effect of information asymmetry

This table reports estimates of the linear regression of FUND_TRADE on ABOVE_VS_BELOW interacted with the following moderating variables: (i) FORECAST_ERROR (Columns 1-2), which represents analyst forecast error, constructed as the absolute value of the difference between analyst forecast of EBIT and the actual EBIT for each firm-year-quarter-end deflated by total assets; (ii) CF_UNCERTAINTY (Columns 3-4), which represents cash flow uncertainty, constructed as the standard deviation of the EBITDA deflated by average total assets from years t-5 to t-1; and (iii) ABS_ACCRUALS (Columns 5-6), measured as the principal component factor—increasing in poor accounting quality—of the absolute value of the residuals from 4 different accruals models estimated in the quarterly setting (i.e., Jones, Modified Jones, DD model and Collins et. al). [See Appendix A for detailed definitions of variables]. Asterisk demarcations *, **, and *** represent statistical significance at 10%, 5%, and 1% levels, respectively. The table reports heteroskedasticity-adjusted t-values in parentheses, observations are clustered at the stock level, and the regression controls for fund and year-quarter FEs. For brevity, estimations here are performed using the maximum bandwidth.

	Linear model					
MODERATING VARIABLE =	FORECAST_ERROR		CF_UNCERTAINTY		ABS_ACCRUALS	
	(1)	(2)	(3)	(4)	(5)	(6)
Cons	-0.114***	-0.081**	-0.020	-0.015	-0.086**	-0.070*
	(-3.29)	(-2.17)	(-0.46)	(-0.32)	(-1.99)	(-1.70)
POSITIVE (unrounded) /100	0.043**	0.033*	0.016	0.007	0.071**	0.054*
	(2.36)	(1.82)	(0.53)	(0.22)	(2.17)	(1.83)
ABOVE_VS_BELOW	0.002	0.002	-0.010	-0.011	-0.009	-0.009
	(0.50)	(0.46)	(-1.27)	(-1.30)	(-0.96)	(-1.09)
ABOVE_VS_BELOW × MODERATING VARIABLE	0.015*	0.016*	0.198*	0.217*	0.022*	0.020*
	(1.80)	(1.94)	(1.68)	(1.76)	(1.93)	(1.86)
POSITIVE (unrounded) × MODERATING VARIABLE	0.034	0.030	0.615*	0.617	0.006	-0.002
	(1.44)	(1.35)	(1.67)	(1.64)	(0.13)	(-0.05)
MODERATING VARIABLE	-0.033*	-0.032*	-0.798**	-0.804**	-0.032	-0.023
	(-1.69)	(-1.71)	(-2.22)	(-2.16)	(-0.77)	(-0.61)
Control Variables	YES	YES	YES	YES	YES	YES
Year-Qtr FE	NO	YES	NO	YES	NO	YES
Fund FE	NO	YES	NO	YES	NO	YES
Cluster by Stock	YES	YES	YES	YES	YES	YES
N	517900	517752	479056	478919	252527	252349
Adjusted R^2	0.007	0.065	0.004	0.031	0.013	0.084

Table 5: Threshold-related heterogeneities

This table reports estimates of a linear regression of FUND_TRADE on ABOVE_VS_BELOW - the indicator for fund trade observations just above vs. below rounding cutoff points [See Appendix A for detailed definitions of variables]. Asterisk demarcations *, **, and *** represent statistical significance at 10%, 5%, and 1% level, respectively. The table reports heteroskedasticity-adjusted t-values in parentheses, observations are clustered at the stock level, and the regression controls for fund and year-quarter FEs. For brevity, estimations here are performed using the maximum bandwidth.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cons	4.081*	4.021*	4.137*	4.076*	-0.103***	-0.075***	-0.105***	-0.078***
	(1.67)	(1.68)	(1.69)	(1.69)	(-4.19)	(-2.86)	(-4.39)	(-3.04)
ABOVE_VS_BELOW	0.020**	0.020**	0.065**	0.061**	0.020**	0.020**	0.020**	0.020**
	(2.41)	(2.41)	(2.02)	(1.98)	(2.41)	(2.41)	(2.38)	(2.37)
POSITIVE (unrounded) /100	-4.999*	-4.894*	-5.066*	-4.959*				
	(-1.71)	(-1.70)	(-1.73)	(-1.72)				
ABOVE_VS_BELOW × POSITIVE (unrounded) /100			-0.054	-0.048				
			(-1.51)	(-1.43)				
RATING_RELATIVE_TO_THRESHOLD					-0.050*	-0.049*	-0.072*	-0.074*
					(-1.71)	(-1.70)	(-1.81)	(-1.96)
ABOVE_VS_BELOW × RATING_RELATIVE_TO_THRESHOLD							0.045	0.052
							(0.63)	(0.75)
Control Variables	YES	YES	YES	YES	YES	YES	YES	YES
Rounding Threshold FE	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**
Year-Qtr FE	NO	YES	NO	YES	NO	YES	NO	YES
Fund FE	NO	YES	NO	YES	NO	YES	NO	YES
Cluster by Stock	YES	YES	YES	YES	YES	YES	YES	YES
N	593030	592889	593030	592889	593030	592889	593030	592889
Adjusted R^2	0.012	0.069	0.013	0.069	0.012	0.069	0.012	0.069

Table 6: Placebo rounding thresholds

This table reports estimates of a linear regression of FUND_TRADE on ABOVE_VS_BELOW - the indicator for fund trade observations just above vs. below rounding cutoff points [See Appendix A for detailed definitions of variables]. Asterisk demarcations *, **, and *** represent statistical significance at 10%, 5%, and 1% levels, respectively. The table reports heteroskedasticity-adjusted t-values in parentheses, observations are clustered at the stock level, and the regression controls for fund and year-quarter FEs. For brevity, estimations here are performed using the maximum bandwidth.

	rounding_thresholds+0.1		rounding_thresholds-0.1		rounding_thresholds+0.15	
	(1)	(2)	(3)	(4)	(5)	(6)
cons	-0.134***	-0.103***	-0.132***	-0.102***	-0.134***	-0.104***
	(-4.10)	(-3.06)	(-3.93)	(-2.92)	(-4.10)	(-3.06)
POSITIVE (unrounded) /100	0.054***	0.049***	0.054***	0.048***	0.055***	0.049***
	(3.06)	(2.74)	(3.04)	(2.72)	(3.06)	(2.77)
ABOVE_VS_BELOW	-0.001	-0.000	-0.001	-0.001	0.003	0.004
	(-0.18)	(-0.04)	(-0.29)	(-0.18)	(0.49)	(0.60)
Control Variables	YES	YES	YES	YES	YES	YES
Year-Qtr FE	NO	YES	NO	YES	NO	YES
Fund FE	NO	YES	NO	YES	NO	YES
Cluster by Stock	YES	YES	YES	YES	YES	YES
N	591862	591721	592925	592784	593030	592889
Adjusted R^2	0.009	0.066	0.009	0.066	0.009	0.066

	rounding_thresholds-0.15		rounding_thresholds+0.2		rounding_thresholds-0.2	
	(7)	(8)	(9)	(10)	(11)	(12)
cons	-0.134***	-0.104***	-0.131***	-0.101***	-0.127***	-0.095***
	(-4.05)	(-3.03)	(-4.05)	(-3.00)	(-3.73)	(-2.64)
POSITIVE (unrounded) /100	0.054***	0.048***	0.052***	0.046**	0.053***	0.047***
	(3.05)	(2.72)	(2.93)	(2.58)	(3.01)	(2.66)
ABOVE_VS_BELOW	0.002	0.002	-0.006	-0.004	-0.006	-0.007
	(0.35)	(0.41)	(-0.68)	(-0.53)	(-0.57)	(-0.67)
Control Variables	YES	YES	YES	YES	YES	YES
Year-Qtr FE	NO	YES	NO	YES	NO	YES
Fund FE	NO	YES	NO	YES	NO	YES
Cluster by Stock	YES	YES	YES	YES	YES	YES
N	593030	593030	592634	592493	592227	592086
Adjusted R^2	0.010	0.009	0.009	0.066	0.009	0.066

Table 7: Actively-managed equity funds vs. passive funds

These tables report estimates of the regression of FUND_TRADE on ABOVE_VS_BELOW - the indicator for fund trade observations just above vs. below rounding cutoff points for exclusive sub-samples of Actively-Managed Equity Funds vs. Passive Funds. [See Appendix A for detailed definitions of variables]. In Panel A, estimates for Actively-Managed Equity (Passive) Funds are reported in Columns 1-2 (3-4). In Panel B, estimates for Actively-Managed Equity (Passive) Funds are reported in Columns 1-4 (5-8). Asterisk demarcations *, **, and *** represent statistical significance at 10%, 5%, and 1% levels, respectively. The table reports heteroskedasticity-adjusted t-values in parentheses, observations are clustered at the stock level, and the regression controls for fund and year-quarter FEs. For brevity, estimations here are performed using the maximum bandwidth.

Panel A: Linear RDD model

	Linear Model			
	Actively-Managed Equity Funds		Passive Funds	
	(1)	(2)	(3)	(4)
Cons	-0.160***	-0.164***	-0.046**	-0.042*
	(-3.55)	(-3.28)	(-2.32)	(-1.95)
POSITIVE (unrounded) /100	0.083***	0.078***	0.027***	0.024***
	(3.24)	(2.98)	(2.84)	(2.67)
ABOVE_VS_BELOW	0.010*	0.011*	0.003	0.003
	(1.75)	(1.75)	(1.33)	(1.32)
Control Variables	YES	YES	YES	YES
Year-Qtr FE	NO	YES	NO	YES
Fund FE	NO	YES	NO	YES
Cluster by Stock	YES	YES	YES	YES
N	221480	221471	196297	196266
Adjusted R^2	0.011	0.040	0.005	0.071

Panel B: Higher order RDD models

	Actively-Managed Equity Funds				Passive Funds			
	Quadratic		Cubic		Quadratic		Cubic	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cons	-0.127**	-0.165**	-0.039	-0.056	-0.019	-0.030	0.002	0.018
	(-2.13)	(-2.57)	(-0.43)	(-0.60)	(-0.65)	(-0.99)	(0.04)	(0.45)
POSITIVE (unrounded) /100	-0.012	0.082	-0.471	-0.483	-0.049	-0.012	-0.158	-0.262
	(-0.10)	(0.67)	(-0.95)	(-0.96)	(-0.83)	(-0.21)	(-0.71)	(-1.26)
square {POSITIVE (unrounded) /100}	0.062	-0.002	0.785	0.886	0.049	0.023	0.221	0.417
	(0.78)	(-0.03)	(1.00)	(1.11)	(1.30)	(0.66)	(0.65)	(1.30)
cubic {POSITIVE (unrounded) /100}			-0.355	-0.436			-0.084	-0.193
			(-0.91)	(-1.10)			(-0.51)	(-1.24)
ABOVE_VS_BELOW	0.010*	0.011*	0.010*	0.011*	0.003	0.003	0.003	0.003
	(1.73)	(1.75)	(1.74)	(1.76)	(1.32)	(1.32)	(1.33)	(1.33)
Control Variables	YES	YES	YES	YES	YES	YES	YES	YES
Year-Qtr FE	NO	YES	NO	YES	NO	YES	NO	YES
Fund FE	NO	YES	NO	YES	NO	YES	NO	YES
Cluster by Stock	YES	YES	YES	YES	YES	YES	YES	YES
N	221480	221471	221480	221471	196297	196266	196297	196266
Adjusted R^2	0.011	0.040	0.011	0.041	0.005	0.071	0.005	0.071

Table 8: Summary statistics (RELIANCE_TWITTER)

Panel A: 1st-stage results for constructing RELIANCE_TWITTER

This table reports the regression coefficient obtained in the 1st stage cross-sectional pooled regression of ΔFUND_TRADE $_q$ on ΔPOSITIVE (unrounded) $_{i,q-1}$. This regression, when estimated (in the spirit of the Kacperczyk and Seru (2007)) at each fund-year-quarter as a cross-sectional regression across stocks (i.e., portfolio firms), derives the measure RELIANCE_TWITTER, which captures the extent of reliance on LikeFolio data and equals the unadjusted R^2 of the regression ΔFUND_TRADE $_q$ on ΔPOSITIVE (unrounded) $_{i,q-1}$.

	ΔFUND_TRADE $_q$
	(1)
_cons	0.000
	(1.12)
ΔPOSITIVE (unrounded) $_{i,q-1}$/100	**0.062*****
	(20.53)
Other controls	**YES**
N	567317
Adjusted R-square	0.002

Panel B: Fund characteristics by quintiles portfolios of RELIANCE_TWITTER

This table reports the variation in TNA (total net assets), dollar expenses, age, and turnover across mutual fund quintiles. We form quintiles every quarter based on the RELIANCE_TWITTER measure and average the variables for the quintiles across all available quarters. RELIANCE_TWITTER measures the extent of reliance on LikeFolio-based Twitter information (customer' satisfaction scores) and equals the unadjusted R^2 of the regression ΔFUND_TRADE $_q$ on ΔPOSITIVE(unrounded) $_{i,q-1}$, constructed similar to the Kacperczyk and Seru (2007) RPI measure. The sample spans the period 2012 to 2015. The last row reports the correlation coefficients of each of the variables with RELIANCE_TWITTER, along with their statistical significance.

Panel C: Quintile Portfolios formed on RELIANCE_TWITTER					
	RELIANCE_TWITTER	TNA	Expense	Age	Turnover
1	0.00	3127.03	0.99	23.73	0.64
2	0.04	3028.10	1.01	23.66	0.68
3	0.13	2350.58	1.05	23.27	0.69
4	0.34	1783.53	1.10	22.40	0.66
5	0.74	1129.53	1.17	22.05	0.67
	1	-0.0735***	0.1564***	-0.0539***	0.0170*

Table 9: Future fund performance and RELIANCE_TWITTER

This table reports the results of the regression of future fund performance (4-factor Carhart alpha $_{f,q+1}$) on RELIANCE_TWITTER while controlling for fund characteristics. RELIANCE_TWITTER measures the extent of reliance on LikeFolio data and equals the unadjusted R^2 of the regression of ΔFUND_TRADE $_q$ on ΔPOSITIVE (unrounded) $_{i,q-1}$, constructed in the spirit of the Kacperczyk and Seru 2007 (RPI) measure. We also construct the reliance on public information (RPI) measure as used in prior studies as the unadjusted R^2 of the regression of Δ*FUND_TRADE* on changes in analysts' last-period consensus recommendation (Kacperczyk and Seru 2007). Fund Size (TNA) is the natural logarithm of total net assets in millions of dollars, Expenses denotes fund expenses, Fund Age is the natural log of one plus the fund age in years which represents the number of years a fund is present in the CRSP equity mutual fund database, Turnover the turnover ratio of a fund, Flow is the growth rate of assets under management after adjusting for the appreciation of the fund's assets (Sirri & Tufano 1998), Team is an indicator variable that equals 1 if a fund is managed by multiple managers and 0 otherwise. Panel B reports the Fama-Macbeth 1973 regression results of future fund performance (3 & 4-factor alpha) on RELIANCE_TWITTER while controlling for fund characteristics. Newey and West (1987) t-statistics are reported in parentheses. In Panel C, RELIANCE_TWITTER (II) represents a measure that captures the reliance on social media (Twitter) and equals the unadjusted R^2 of the regression of percentage changes in fund managers' portfolio holdings on ABOVE_VS_BELOW (which represents an exogenous 1-percentage [point] change in customer satisfaction/happiness scores), constructed in the spirit of Kacperczyk and Seru 2007 RPI. We include Time, Fund family, Style and Manager's FEs and we also cluster by Fund and Time.

	Panel A: Rate of adoption based on ΔPOSITIVE (unrounded)				
	Dep. var. = 4-factor alpha $_{f,q+1}$				
	(1)	(2)	(3)	(4)	(5)
Cons	-0.130***	-0.155***	-0.105*	-0.161	-0.130**
	(-3.41)	(-2.59)	(-1.76)	(-1.18)	(-2.77)
RELIANCE_TWITTER	0.038***	0.029***	0.019**	0.020***	0.038***
	(3.86)	(3.47)	(2.29)	(2.97)	(2.55)
RPI	-0.002	-0.006	-0.007	-0.006	-0.002
	(-0.33)	(-0.95)	(-1.14)	(-1.05)	(-0.33)
Fund Size (TNA)	0.020***	0.018***	0.018***	0.013*	0.020***
	(6.90)	(4.45)	(4.36)	(1.87)	(6.50)
Fund Age	-0.005	-0.014	-0.022	-0.004	-0.005
	(-0.44)	(-0.86)	(-1.28)	(-0.11)	(-0.40)
Expenses	-0.071***	-0.002	-0.028	-0.023	-0.071***
	(-4.66)	(-0.08)	(-1.07)	(-0.54)	(-4.56)
Turnover	-0.013	-0.031***	-0.029***	-0.023***	-0.013
	(-1.03)	(-2.86)	(-2.78)	(-2.59)	(-0.95)
Flow	0.265***	0.186***	0.188***	0.137***	0.265***
	(7.86)	(7.53)	(7.58)	(6.66)	(6.15)
Team	0.004	0.012	0.013	0.049**	0.004
	(0.39)	(0.78)	(0.85)	(2.05)	(0.40)
Time FE	YES	YES	YES	YES	YES
Family FE	NO	YES	YES	YES	NO
Style FE	NO	NO	YES	YES	NO
Manager FE	NO	NO	NO	YES	NO
Cluster by Fund	YES	YES	YES	YES	YES
Cluster by Time	NO	NO	NO	NO	YES
N	11258	11258	11258	11258	11258
Adjusted R^2	0.135	0.369	0.381	0.541	0.135

Panel B: Fama-Macbeth Rate of adoption based on ΔPOSITIVE (unrounded)	
	Dep. var: = 4-factor alpha $_{f,q+1}$
RELIANCE_TWITTER	0.040***
	(3.65)
Control Variables	YES
N	11258
Adjusted R^2	0.127

	Panel C: Rate of adoption based on ABOVE_VS_BELOW				
	Dep. var: = 4-factor alpha $_{f,q+1}$				
	(1)	(2)	(3)	(4)	(5)
Cons	-0.130***	-0.155***	-0.105*	-0.161	-0.130**
	(-3.41)	(-2.59)	(-1.76)	(-1.18)	(-2.77)
RELIANCE_TWITTER (II)	0.032***	0.026***	0.019**	0.013*	0.032***
	(3.34)	(3.32)	(2.48)	(1.86)	(2.72)
Control Variables	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Family FE	NO	YES	YES	YES	NO
Style FE	NO	NO	YES	YES	NO
Manager FE	NO	NO	NO	YES	NO
Cluster by Fund	YES	YES	YES	YES	YES
Cluster by Time	NO	NO	NO	NO	YES
N	11258	11258	11258	11258	11258
Adjusted R^2	0.134	0.369	0.381	0.540	0.134

Table 10: Holding-based measures—DGTW returns and RELIANCE_TWITTER

This table reports the results of the regression of future holding-based performance measure by Daniel, Grinblatt, Titman and Wermers (1997) on RELIANCE_TWITTER while controlling for fund characteristics. RELIANCE_TWITTER measures the extent of reliance on LikeFolio data and is constructed similar to the Kacperczyk and Seru (2007) RPI measure. Fund Size (TNA) is the natural logarithm of total net assets, Expenses denotes fund expenses, Fund Age is the natural logarithm of fund age, Turnover is the fund's turnover. We include Time, Fund family, Style and Manager's Fixed effects and we also cluster by Fund and Time.

	Panel A: Characteristic Selectivity (CS)				
	(1)	(2)	(3)	(4)	(5)
Cons	-0.141**	-0.346***	-0.251***	-0.188	-0.141
	(-2.19)	(-3.68)	(-2.61)	(-0.85)	(-0.72)
RELIANCE_TWITTER	0.080***	0.092***	0.073**	0.083**	0.080
	(2.66)	(2.80)	(2.20)	(2.22)	(1.40)
Control Variables	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Family FE	NO	YES	YES	YES	NO
Style FE	NO	NO	YES	YES	NO
Manager FE	NO	NO	NO	YES	NO
Cluster by Fund	YES	YES	YES	YES	YES
Cluster by Time	NO	NO	NO	NO	YES
N	11258	11258	11258	11258	11258
Adjusted R^2	0.106	0.121	0.122	0.114	0.106

	Panel B: Characteristic Timing (CT)				
	(1)	(2)	(3)	(4)	(5)
Cons	0.068**	0.023	0.027	0.246*	0.068**
	(2.29)	(0.41)	(0.45)	(1.79)	(2.28)
RELIANCE_TWITTER	0.027**	0.026*	0.025*	0.017	0.027
	(2.05)	(1.84)	(1.69)	(1.16)	(0.87)
Control Variables	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Family FE	NO	YES	YES	YES	NO
Style FE	NO	NO	YES	YES	NO
Manager FE	NO	NO	NO	YES	NO
Cluster by Fund	YES	YES	YES	YES	YES
Cluster by Time	NO	NO	NO	NO	YES
N	11258	11258	11258	11258	11258
Adjusted R^2	0.020	0.045	0.046	0.065	0.020

	Panel C: Average Style (AS)				
	(1)	(2)	(3)	(4)	(5)
Cons	-0.141**	-0.360***	-0.121	-0.456**	-0.141
	(-2.12)	(-3.01)	(-1.12)	(-1.99)	(-0.67)
RELIANCE_TWITTER	0.084***	0.048	-0.004	-0.010	0.084
	(2.67)	(1.35)	(-0.10)	(-0.26)	(0.54)
Control Variables	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Family FE	NO	YES	YES	YES	NO
Style FE	NO	NO	YES	YES	NO
Manager FE	NO	NO	NO	YES	NO
Cluster by Fund	YES	YES	YES	YES	YES
Cluster by Time	NO	NO	NO	NO	YES
N	11258	11258	11258	11258	11258
Adjusted R^2	0.845	0.845	0.846	0.844	0.845

Table 11: Future fund performance and RELIANCE_TWITTER (dynamic regression)

This table reports regressions of future fund performance (4-factor alpha) on RELIANCE_TWITTER interacted with year-quarter dummies while controlling for fund characteristics. RELIANCE_TWITTER measures the reliance on twitter information (stocks' consumer satisfaction scores) and is constructed similar to the Kacperczyk and Seru (2007) RPI measure. We include all control variables from the baseline. We include Fund, Time, Fund family, Style and Manager's FEs and we also cluster by Fund. "Sub-sample" in the table below indicates scenarios where RELIANCE_TWITTER is greater than 0.5.

	4 factor alpha	
	(1)	(2)
cons	-0.191	-0.237**
	(-1.61)	(-2.56)
RELIANCE_TWITTER	0.021	0.013
	(1.20)	(0.77)
RELIANCE_TWITTER × 2012Q1	0.000	0.000
	(.)	(.)
RELIANCE_TWITTER × 2012Q2	-0.017	-0.013
	(-0.69)	(-0.55)
RELIANCE_TWITTER × 2012Q3	-0.000	0.004
	(-0.00)	(0.20)
RELIANCE_TWITTER × 2012Q4	0.061***	0.070***
	(2.80)	(3.29)
RELIANCE_TWITTER × 2013Q1	0.074***	0.084***
	(3.54)	(4.05)
RELIANCE_TWITTER × 2013Q2	0.052**	0.045**
	(2.50)	(2.23)
RELIANCE_TWITTER × 2013Q3	0.002	0.007
	(0.10)	(0.36)
RELIANCE_TWITTER × 2013Q4	0.021	0.014
	(1.03)	(0.72)
RELIANCE_TWITTER × 2014Q1	-0.016	-0.010
	(-0.86)	(-0.52)
RELIANCE_TWITTER × 2014Q2	0.003	0.008
	(0.15)	(0.37)
RELIANCE_TWITTER × 2014Q3	-0.006	-0.004
	(-0.20)	(-0.14)
RELIANCE_TWITTER × 2014Q4	-0.046	-0.033
	(-1.64)	(-1.19)
RELIANCE_TWITTER × 2015Q1	-0.020	-0.015
	(-0.89)	(-0.66)
RELIANCE_TWITTER × 2015Q2	-0.016	-0.006
	(-0.53)	(-0.21)
RELIANCE_TWITTER × 2015Q3	-0.047	-0.038
	(-1.47)	(-1.15)
RELIANCE_TWITTER × 2015Q4	-0.082***	-0.077**
	(-2.57)	(-2.44)
Control Variables	YES	YES
Time FE	YES	YES
Family FE	YES	YES
Style FE	YES	YES
Manager FE	YES	YES
Fund FE	NO	YES
Cluster by Fund	YES	YES
N	11258	11258
Adjusted R^2	0.511	0.523

Table 12: Future cumulative abnormal returns and RELIANCE_TWITTER (dynamic regression)

This table reports regression estimates of 90-days-ahead cumulative equally-weighted abnormal returns (at each firm-year-quarter-end) CAR $_{i,q+1}$ on the interaction between RELIANCE_TWITTER_S, ABOVE_VS_BELOW and year-quarter dummies. RELIANCE_TWITTER_S is a dummy equal to 1 if the firm's rate of adoption of LikeFolio data is above the median and 0 otherwise. A firm's rate of adoption of LikeFolio data is the weighted average of all funds' RELIANCE_TWITTER, conditional on the fact that the firm is covered in the portfolio of these funds. [See Appendix A for detailed definitions of variables]. Asterisk demarcations *, ** and *** represent statistical significance at 10%, 5% and 1% levels, respectively. The table reports heteroskedasticity-adjusted t-values in parentheses, observations are clustered at the stock level and the regression controls for firm, year-quarter and industry FEs.

		CAR $_{i,q+1}$
	(1)	(2)
Cons	0.088	0.323
	(0.25)	(0.87)
RELIANCE_TWITTER_S	-0.348	-0.099
	(-1.40)	(-0.40)
ABOVE_VS_BELOW	-0.001	0.010
	(-0.03)	(0.25)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW	-0.045	-0.057
	(-0.53)	(-0.58)
POSITIVE (unrounded)	-0.144	0.182
	(-0.45)	(0.57)
RELIANCE_TWITTER_S × POSITIVE (unrounded)	0.299	-0.088
	(0.84)	(-0.26)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2012Q1	0.000	0.000
	(.)	(.)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2012Q2	0.014	0.046
	(0.08)	(0.24)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2012Q3	0.177	0.220
	(1.24)	(1.58)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2012Q4	**0.275****	**0.306***
	(2.01)	(1.91)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2013Q1	**0.194**	**0.028**
	(1.22)	(0.11)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2013Q2	**0.192**	**0.195**
	(1.11)	(1.06)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2013Q3	0.094	0.061
	(0.83)	(0.51)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2013Q4	0.068	0.030
	(0.60)	(0.25)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2014Q1	-0.194	-0.314
	(-0.55)	(-1.53)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2014Q2	0.079	0.107
	(0.63)	(0.79)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2014Q3	-0.008	-0.022
	(-0.03)	(-0.09)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2014Q4	0.028	0.034
	(0.15)	(0.18)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2015Q1	-0.027	-0.179
	(-0.04)	(-0.24)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2015Q2	-0.105	-0.089
	(-0.95)	(-0.75)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2015Q3	0.065	0.049
	(0.66)	(0.43)
RELIANCE_TWITTER_S × ABOVE_VS_BELOW × 2015Q4	0.044	-0.007
	(0.51)	(-0.06)
RELIANCE_TWITTER_S × Year-Qtr. interactions	YES	YES
ABOVE_VS_BELOW × Year-Qtr. interactions	YES	YES

RELIANCE_TWITTER_S × POSITIVE (unrounded) × Year-Qtr. Interactions	YES	YES
POSITIVE (unrounded) × Year-Qtr. interactions	YES	YES
Control Variables	NO	YES
Year-Qtr FE	YES	YES
Industry FE	YES	YES
Cluster by Stock	YES	YES
N	1568	1568
Adjusted R^2	0.429	0.434

ONLINE APPENDIX

Table OA1: Higher Frequency Test—daily trading volume and daily product tweets

This table reports the regression of the natural log of daily CRSP trading volume (LOG_VOL) on ABOVE_VS_BELOW –the indicator for rounding up vs. down, a 30-day moving average of Twitter-based consumer satisfaction scores—POSITIVE (unrounded). A 30-day moving average of daily sentiments is used as the proxy for POSITIVE (unrounded) because this is one of the 3 options (7-day vs. 30-day vs. 90-day moving average) that LikeFolio displays daily to users (in rounded form) on the dashboard of their subscription-based app (NB: the free app on the other hand displays only the 30-day moving average) [see figures 1A and 1B]. [See Appendix A for detailed definitions of variables]. Asterisk demarcations *, ** and *** represent statistical significance at 10%, 5% and 1% levels, respectively. The table reports heteroskedasticity-adjusted t-values in parentheses, observations are clustered at the stock level and the regression controls for alternative FEs. Because quite a large number of observations are missing for the control variables, to preserve test power when using LOG_VOL as the dependent variable, we adopt a common approach for dealing with missing observations and fill in missing values with the industry-date mean, while simultaneously creating a corresponding missing value indicator which we include alongside all control variables in regressions. The missing value indicator, which we specify in regressions, is expected to control for any measurement error associated with the adjustment we had made to each control variable.

	(1)	(2)	(3)	(4)	(5)
			LOG_VOL		
_CONS	8.919***	13.939***	12.046***	13.789***	13.966***
	(20.72)	(50.52)	(23.02)	(185.98)	(126.52)
ABOVE_VS_BELOW	**0.036*** **	**0.006***	**0.008***	**0.006***	**0.009***
	(3.36)	(1.72)	(1.66)	(1.85)	(1.81)
POSITIVE (unrounded) /100	-1.156***	0.041	0.096*	0.177**	0.263*
	(-5.43)	(0.97)	(1.78)	(1.96)	(1.95)
SIZE_ASSETS	0.712***	0.004	0.236***		
	(21.89)	(0.13)	(4.39)		
PROFITABILITY	-0.084	-0.098	-0.156		
	(-0.46)	(-1.12)	(-1.56)		
TANGIBILITY_NET	0.432*	-0.014	-0.107		
	(1.91)	(-0.05)	(-0.29)		
LEVERAGE	-0.187	0.126	0.077		
	(-1.18)	(1.10)	(0.63)		
MARKET_TO_BOOK	0.097***	0.013*	0.027***		
	(3.92)	(1.70)	(2.93)		
Z_SCORE	-0.051***	0.012**	0.014**		
	(-3.68)	(2.19)	(1.98)		
SALES_GROWTH	0.014	-0.026	-0.042		
	(0.29)	(-0.94)	(-1.36)		
INVENTORY	1.073**	-0.465	0.522		
	(2.45)	(-1.15)	(1.00)		
LOSS	0.223**	0.004	-0.015		
	(2.03)	(0.16)	(-0.63)		
OPERATING_CYCLE	-0.113**	-0.018	-0.022		
	(-2.00)	(-0.76)	(-0.90)		
CASH_HOLDINGS	2.071***	0.054	-0.045		
	(6.31)	(0.35)	(-0.24)		
DIVIDEND	-0.185	0.021	0.015		
_	(-1.53)	(0.44)	(0.30)		
Firm FE	No	Yes	Yes	Yes	Yes
Date FE (i.e., Year-Month-Day FE)	No	Yes	No	No	No
Industry*Date FE	No	No	Yes	No	Yes
Firm-Year-Month-Week FE	No	No	No	Yes	Yes
N	182202	182182	99217	163979	87144
Adjusted R^2	0.282	0.899	0.907	0.943	0.944

Table OA2: OLS model

Fund Trade and exact (unrounded) consumer satisfaction scores

This table reports estimates of a OLS regression of FUND_TRADE on exact (unrounded) customer satisfaction scores [See Appendix A for detailed definitions of variables]. Asterisk demarcations *, **, and *** represent statistical significance at 10%, 5%, and 1% levels, respectively. The table reports heteroskedasticity-adjusted t-values in parentheses, observations are clustered at the stock level, and the regression controls for fund and year-quarter FEs.

	Panel A: Fund Trade (Max bandwidth)	
	(1)	(2)
Cons	-0.155***	-0.095***
	(-5.98)	(-4.45)
POSITIVE (unrounded) /100	**0.038*** **	**0.027** **
	(2.76)	**(2.25)**
SIZE_ASSETS	0.018***	0.013***
	(8.12)	(6.45)
PROFITABILITY	0.052	0.051
	(0.90)	(0.99)
TANGIBILITY_NET	-0.008	-0.009
	(-0.75)	(-1.16)
LEVERAGE	-0.024***	-0.013**
	(-2.76)	(-2.32)
MARKET_TO_BOOK	0.009***	0.008***
	(6.14)	(5.96)
Z_SCORE	-0.001	-0.001
	(-0.47)	(-0.61)
SALES_GROWTH	-0.003	-0.003
	(-0.50)	(-0.55)
INVENTORY	0.007	0.021**
	(0.52)	(2.14)
LOSS	0.002	-0.004
	(0.52)	(-1.12)
OPERATING_CYCLE	-0.001	-0.004*
	(-0.42)	(-1.88)
CASH_HOLDINGS	0.027	0.020
	(1.46)	(1.24)
DIVIDEND	-0.022***	-0.012**
	(-3.40)	(-2.29)
Year-Qtr FE	NO	YES
Fund FE	NO	YES
Cluster by Stock	YES	YES
N	1327439	1327362
Adjusted R^2	0.004	0.220