Tech-Enabled Financial Data Access, Retail Investors, and Gambling-like Behavior in the Stock Market: Evidence from a Natural Experiment

Taha Havakhor
Temple University
taha.havakhor@temple.edu

Mohammad S. Rahman
Purdue University
mrahman@purdue.edu

Tianjian Zhang
City University of Hong Kong
tj.zhang@cityu.edu.hk

Chenqi Zhu*
University of California, Irvine
chenqiz1@uci.edu

Abstract

A significant portion of retail investors heavily engage in feedback trading, which is built on historical price data. As financial technologies lower individuals' acquisition cost to these data, the abundance of (raw) information creates an illusion of knowledge for retail investors and boosts their overconfidence, which further induces them to trade too much. Against this backdrop, we investigate the impact of technology-enabled convenient access to financial data on retail investments. Our identification strategy exploits the sudden shutdown of Yahoo! Finance Application Programming Interface (API), which cut off the largest free price data access for retail investors engaging in feedback trading. We find that within one month after the API shutdown, retail trading volumes in stocks favored by those investors dropped by 8.6%-10.5%. The remaining retail trades became more predictive of future returns, suggesting less gambling-like behavior after the API shutdown. The study reveals a dark side of technology-led wider data provision to retail investors, and echoes regulators' call to improve the financial literacy of retail investors.

Keywords: Retail investors, financial technology, gambling, noise trading, financial market, application programming interface, natural experiment

JEL Codes: G12, G14, O33, H41

* The authors thank Gregory Eaton, Keongtae Kim, Soyoon Kong, Min-Seok Pang, Ben Liu, seminar and conference participants at the City University of Hong Kong, Georgia State University, 2020 Statistical Challenges in Electronic Commerce Research, 2019 Conference on Information Systems and Technology, 2019 Conference on Digital Experimentation, and 2019 Workshop on Information Systems and Economics. All errors are ours.

1. Introduction

Retail investors - individuals who invest in the stock market for their personal accounts - their well-being and their potential impact on financial markets have long been of interest to economists, policy-makers, and regulators in the US. Specifically, the increasingly convenient access of retail investors to financial technology has led to a series of dramatic events[1, 2] that have further captured the attention of those concerned parties, posing questions and sparking hot debates about how these technologies impact the main-street traders. Existing financial technologies include a broad range, from platforms helping rapid diffusion of knowledge and information about trading, such as *Reddit* and *Seeking Alpha*, to applications facilitating low-(transaction)cost trading for households, e.g., *Robinhood*. While these technologies can level the playing field for retail investors, they also give rise to vast uncertainties.

Notably, financial data application programming interfaces (APIs) constitute an overwhelmingly understudied category of financial technologies that have enabled easy access to a huge amount of raw (often historical) market data, upon which individuals can devise their own trading strategies. However, individuals typically lack complementary capabilities (e.g. financial literacy) to properly process and interpret technology-enabled raw data, as opposed to directly taking advice from investment research on platforms such as Seeking Alpha. As raw market data (marginally) improve the information environment for the "small guys," this very potential also entices inexperienced investors to trade excessively (Han et al. 2016; Zhang and Zhang 2015). Against this backdrop, this study examines how financial technologies that facilitate raw market data dissemination affect retail investments.

Retail investors are often dubbed as noise traders. This is mainly due to two reasons – they lack information access, plus various behavioral biases and limited financial illiteracy make them ignore useful information. While financial technology can improve the information environment for retail investors, it does not eradicate their cognitive biases. Many retail investors focus on past price trends and ignore key

[1] https://www.forbes.com/sites/sergeiklebnikov/2020/06/17/20-year-old-robinhood-customer-dies-by-suicide-after-seeing-a-730000-negative-balance/?sh=484d2e301638
[2] https://www.vox.com/the-goods/22249458/gamestop-stock-wallstreetbets-reddit-citron

accounting information (Blankespoor et al. 2019). As such, these individuals often trade impulsively, favor attention-grabbing stocks, and are overly confident (Barber and Odean 2007; Grinblatt and Keloharju 2009). These investors even treat stock market investments like gambling (Dorn et al. 2015; Gao and Lin 2015; Kumar 2009). For this type of retail investor, sorting through a sea of historical stock prices looking for patterns may create a false sense of knowledge and control. The exacerbated overconfidence will lead to more excessive trading. In this case, not only does financial technology not fix behavioral biases, it aggravates them.

Our empirical strategy exploits a natural experiment - the abrupt shutdown of the Yahoo! Finance API, the most popular financial API at the time. On May 16, 2017, Yahoo! Finance API was suddenly shut down without any warnings or announcements. Yahoo! Finance is popular among retail investors (Lawrence et al. 2016). Since its financial API mainly provides historical stock prices, it is more appealing to retail investors actively engaging in feedback trading. These retail investors trade based on the trend of the market and often ignore fundamental information (Blankespoor et al. 2019). Due to their behavioral heuristics, their investments often congregate on stocks like "gambling stocks" (stocks that offer a skewed payoff) as they view trading as a get-rich-quick scheme and treat stocks like lotteries (Dorn et al. 2015; Gao and Lin 2015). Therefore, the shutdown disproportionately affects this group of retail investors. Leveraging this effect, we employ a difference-in-difference (DID) design. While obtaining large-scale information on the actual consumption of market data by retail investors is difficult, we try to understand the extent of raw data consumption by retail investors by comparing retail trades in retail-favored stocks (treatment group) with retail trades in other stocks (control group) before and after the API shutdown.

We find that in a two-month window centered around the shutdown of Yahoo! Finance API, the retail trading volume drops by approximately 8.6%-10.5% in retail-favored stocks. In the meantime, the remaining retail trades become better predictors for future returns, suggesting more informed trading and less feedback trading following the API shutdown. In other words, the API-enabled quick and convenient access to raw financial data increases gambling-like behaviors by retail investors. When we extend the window to four months, the decrease in retail trading volumes and the increase in return predictability become much smaller and statistically insignificant, suggesting that gambling-like behaviors return as retail investors switched to alternative sources of market data (e.g., other APIs).

Additionally, as retail investors are vital market participants and important liquidity providers (Grossman and Miller 1988; Kaniel et al. 2008; Kelley and Tetlock 2013), their absence will negatively affect market liquidity. Indeed, we find that the Amihud illiquidity measure (price impact per share traded) and bid-ask spread increase by 12.3%-17.8% and 5.1%-7.9%, respectively, in the month after the shutdown of Yahoo! Finance API for retail-favored stocks, relative to other stocks. This deterioration in overall market liquidity following the shutdown is consistent with public information improving market liquidity by attracting uninformed trading (Han et al. 2016).

We conduct two tests to rule out alternative stories. To address the concern that our results are driven by confounding events, as a placebo test, we examine institutional trading volume around the Yahoo! Finance API shutdown. If it were confounding events that affect market conditions that further reduce retail trades, we should also observe a reduction in institutional trades. However, we do not find any significant change. To alleviate confounders such as retail trading seasonality, we repeat our main analyses in a two-month window around May 16, 2016, one year before the actual shutdown of Yahoo! Finance API. Again, we do not find any significant changes in retail trades in this 2016 sample, thus undercutting the plausibility of retail trading seasonality and providing further support to our conclusion.

This study adds to the literature on the role of financial technologies in retail investments. While prior studies suggest technologies that produce or disseminate processed information or fundamental information improves retail trades (Farrell et al. 2018; Gao and Huang 2020), we show that technologies disseminating raw market data aggravates excessive trading by retail investors and to their detriment. Therefore, it is important to distinguish the type of information enabled by the technology and how it interacts with retail investors' behavioral biases. The study also provides additional evidence where a seemingly beneficial financial technology can produce unintended consequences for retail investors (Barber and Odean 2002; Zhang and Zhang 2015).

Our study aligns with SEC's call to improve retail investors' financial literacy (Stein 2018). In the United States, retail investors are the biggest owners of stocks - greater than institutions such as mutual funds, pension funds, and hedge funds[3]. Despite their size and importance in providing market liquidity, retail trades often perform poorly (Barber and Odean 2000). This is because their lack of financial knowledge coupled with cognitive biases make them trade excessively and aggressively. As our study shows that financial technology can exacerbate excessive retail trading, what retail investors urgently need is financial knowledge, not necessarily access to more (raw) data. Providing them with the necessary financial training can help retail investors understand the risk in following historical prices, and prevent them from trading too frequently or aggressively.

2. Background

2.1. Yahoo! Finance API

Yahoo! Finance is the most popular website for financial information, attracting over 30 million unique daily users (Lawrence et al. 2016). Yahoo's extensively built database of market transactions over the years and a free-access model make Yahoo! Finance API a popular tool among small investors. Instead of paying $20,000 plus annual subscription fees for a Bloomberg terminal, individuals can conveniently access a wealth of historical price data for free through the API. While the exact number of the API users is not publicly known, its popularity can be inferred from web search volumes. Based on Google search volumes, the Yahoo! Finance API was the most searched compared with other alternatives (Appendix A).

The key value of Yahoo! Finance API for retail investors is the efficient access to a large amount of historical market data (e.g., open and close price), and basic financial variables such as fifty days' moving average (see Appendix B for an extended list of variables accessible through the API). These data are critical for investors conducting feedback trading, which relies extensively on tracking historical price trends. Individuals can connect their data analytics software such as VBA, MATLAB, and R with Yahoo! Finance API for their own analyses. Platforms such as Youtube, Stack Overflow, Quora provide various tutorials

[3] https://www.nasdaq.com/articles/what-everyone-should-know-about-the-stock-market-2020-09-10

and cues on how to access the historical data using different software (Appendix C). These platform interactions also elucidate a community of interest in accessing historical prices through APIs. As one retail investor mentioned in the Yahoo! Finance Help Community[4]:

> *"For over six years I have been using http:/ichart.finance.yahoo.com for downloading historical data programmatically using an interpreter written in Java and it has been a very good experience."*

Additionally, some applications facilitating feedback trading and technical analyses (e.g. Amibroker[5]) relied on the API for historical data. As such, the API functioned as a critical gateway for retail investors relying heavily on historical price data.

On May 16, 2017, Yahoo! Finance API was abruptly shut down. The access to historical data was completely wiped out. Notably, visitors to the Yahoo! Finance webpage could still manually download the spreadsheet files to access historical data. However, this is extremely tedious if one intends to access the data in bulks (Appendix D). While there are a few alternative APIs that provide similar functionalities, their contents are either more costly or less extensive. Instant switching to different APIs is also unlikely given the learning curves and the fact that some users initially were waiting for the Yahoo API to go back online. Essentially, the API shutdown terminated the means to download free historical price data in bulks.

While Yahoo! did not offer any notice before or any explanations after the shutdown[6], it was speculated, among many others, that the main reason for the shutdown was due to financial concerns and potential term violation for third-party data redistribution[7]. At any rate, the shutdown caught its users by surprise, as evidenced by the outrage in a thread with over 250 replies within the Yahoo! Finance Help Community. Even the Yahoo! Finance Help Community administrator was caught off guard (Appendix D).

Taken together, the sizable userbase and the abrupt nature of the API shutdown provide us with a plausibly exogenous natural intervention to understand how such financial technologies affect the behavior of a large group of retail investors.

---

[4] https://web.archive.org/web/20170828230516/https://forums.yahoo.net/t5/Yahoo-Finance-help/Is-Yahoo-Finance-API-broken/td-p/250503/page/3

[5] https://www.amibroker.com/

[6] Even Altaba's SEC filings in the subsequent period do not make any mention to the shutdown of the API.

[7] http://blog.intrinio.com/yahoo-finance-api-replacement/

2.2. Retail Investors

The behavior of retail investors is a topic of broad interest in the Finance and Accounting literature. Like other investors, there are enormous heterogeneities among retail investors in terms of financial sophistication and behavioral biases (e.g., see Barber and Odean (2013) for a recent review). On the one end of the spectrum, some retail investors prefer passive trading strategies (e.g., buy and hold value stocks or copy other's trading ideas (Apesteguia et al. 2020; Von Gaudecker 2015)) and make more rational investment decisions (Dhar and Zhu 2006). On the other end, we see active day traders treat trading like gambling and bet on stocks without solid fundamental analyses of the investment targets (Grinblatt and Keloharju 2009). Such retail investors (hereafter referred to as *active retail investors*) are the focus of the finance and accounting literature, hence ours as well.

Active retail investors are characterized as traders with limited financial knowledge hence heavy reliance on technical trend analyses as opposed to fundamental analyses (Blankespoor et al. 2019; Zhang and Zhang 2015), constrained access to information (Blankespoor et al. 2018; Gao and Huang 2020), higher susceptibility to behavioral heuristics such as limited attention, sensation seeking, and overconfidence (Barber and Odean 2001; Barber and Odean 2007; Grinblatt and Keloharju 2009). Their limited information access and processing capability, augmented with overconfidence, make these retail investors more vulnerable to the gambling instincts, luring them into a trading strategy that is hazardous to their wealth: trading too often and trading on stocks that do not match their risk tolerance. Barber and Odean (2000) show that on average, retail investors lose money compared to just holding the index fund. The more they trade, the more money they lose. Kumar (2009) and Gao and Lin (2015) find that the overconfidence-induced gambling-like trading concentrates on lottery-like stocks with large volatility, high skewness, a "cheap" price tag. A case in point is the recent saga of AMC and GameStop.[8]

One would think that providing active retail investors with easier access to information would relax one limitation they face (i.e., ease the informational friction) hence improve their investment decision making,

[8] https://www.forbes.com/advisor/investing/gamestop-meme-stocks-bb-amc-nok/

at least not worse off, ceteris paribus. Indeed, Gao and Huang (2020) find that retail investors make more informative trades after the implementation of the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system by the SEC which distributes fundamental information (e.g., annual reports) on the internet for free. Since Yahoo! Finance API also streamlines the speedy acquisition of fundamental information such as earnings per share and price-to-earnings ratios on a large scale, we expect that some retail investors could use such information to make better trades.

However, it is unclear whether active retail investors will put the fundamental information in API to good use for two reasons: active retail investors mostly conduct technical analyses and feedback trading built on past stock prices (Blankespoor et al. 2019; Zhang and Zhang 2015) and they may not have sufficient financial acumen to conduct fundamental analyses (Lusardi and Mitchell 2014). Indeed, Blankespoor et al. (2019) find that even when fundamental information is as readily available as historical price information, retail investors rely excessively on historical price information due to their limited financial knowledge and high information processing costs (Blankespoor et al. 2019).

Moreover, since the dominant trading strategy among active retail investors is built on past stock prices, Yahoo API is more likely to influence their trading propensity and quality by enabling them to access a large scale of stock price information quickly and conveniently. As mentioned in Section 2.1, the vast majority of the information available through the Yahoo API is indeed price data. Appendix C provides illustrative examples that these retail investors actively seek knowledge on how to access and interpret historical return data in popular crowd-wisdom sources such as YouTube, Stack Overflow, and Quora. Wherein armed with historical data facilitated by Yahoo API, but often relying on flawed trading strategies (relying on past trends do not generate positive returns on average (Da et al. 2021)), these active investors could make worse investment decisions.

Furthermore, additional information could magnify the influence of other limitations such as boosting retail investors' overconfidence through the illusion of knowledge and perceived self-efficacy (Barber and Odean 2002). Specifically, Zhang and Zhang (2015) show that with IT infrastructure making trading more convenient and information more readily available, uninformed investors adopt more aggressive trading strategies which create higher risks without higher benefits. Similarly, Han et al (2016) find that more public information crowds out private information acquisition and induces uninformed investors to trade more to their detriment. After having ensembled a large amount of data through the API and playing with the data, active retail investors are tempted to believe that such data confers knowledge and develop the illusion that they know better than the average investor in the market and their trades will beat others'. Perceiving themselves to be more competent, they will trade more often (Graham et al. 2009). The excessive trades induced by the elevated confidence level and inflated knowledge illusion are not as good as existing trades, leading to a deterioration in the average quality of their trades.

Taken together, financial APIs are more appealing to active retail investors who rely on technical analyses, trade stocks like gambling, and are vulnerable to overconfidence, as APIs provide easy and fast access to a large amount of historical price data to enable technical analyses, identify lottery-like stocks, and create the illusion of knowledge to boost over-confidence. Therefore, we expect that financial APIs that provide historical price data are more likely to affect active retail investors.

3. Empirical Design

3.1 Identification

The abrupt shutdown of Yahoo Finance API, a major source of free historical price data, provides us a natural experiment to study the impact of financial technologies on retail investors. As discussed in Section 2, active retail investors who conduct technical analysis and feedback trading are more likely to be consumers of Yahoo Finance API and, therefore, are disproportionately affected by its shutdown. While we do not have investor-level data, active retail investors' trades can be inferred from the stocks that they prefer to invest in. Specifically, we leverage the well-established *clientele effect*, which suggests "*different investors restrict themselves to trading within different natural 'habitats' or groups of stocks*" (Kumar and Lee 2006, P. 2452) due to their preference or limitation (Barber and Odean 2000; Conrad et al. 2014; Gao and Lin 2015; Grinblatt and Keloharju 2009; Ivković et al. 2008). For example, due to sensation-seeking preference or get-rich-quick desire, active retail investors prefer stocks that exhibit lottery-like payoffs, (Kumar 2009). It has also been documented that retail investors actively trade on small-cap stocks, attention-grabbing stocks, or value stocks (Barber and Odean 2000; Barber and Odean 2007; Kumar and Lee 2006). Therefore, we construct the treatment group as stocks favored by active retail investors and the control group consists of the remaining stocks. This allows us to examine retail trades in stocks favored by active retail traders relative to retail trades in other stocks before and after the API shutdown.

We use various proxies for the revealed preferences of active retail investors. First, following the spirit of Gao and Lin (2015), Sias and Starks (1997), and Kumar and Lee (2006), we identify *retail-favored stocks* as those held primarily by retail investors. The ownership structure is a snapshot that uncovers the revealed preference of retail investors. Retail holdings are measured as shares not reported to be held by institutional investors, scaled by total shares outstanding. We construct a dummy variable *RFS*, which takes the value of one when retail holding is above the sample median, to indicate retail-favored stocks. This provides a natural distribution of stocks (RFS = 1) whose retail trades are mostly impacted by the API and other stocks whose retail trading is largely unimpacted (i.e., control firms; RFS = 0). In other words, retail-favored stocks are the treatment group while other stocks fall into the control group.

We also construct alternative proxies to identify stocks preferred by active retail investors. As active retail investors treat trading as gambling to some extent and like to bet on lottery-like stocks (Grinblatt and Keloharju 2009), we construct a lottery-likeness index (*Lottery_Like*) which is equal to one when the stock has above-median volatility (*High_Volatility*), above-median skewness (*High_Skewness*), and below-median price (*Low_Priced*), following Kumar (2009).[9] The idea is that gambling-motivated trading concentrates on stocks whose payoff structure resembles a lottery that offers a cheap bet with a small probability to generate extreme positive returns. For completeness, we also use volatility, skewness, and price level as a single defining characteristic to identify stocks that appear like a lottery (Appendix F). As

[9] Following Kumar (2009), volatility is defined as the standard deviation of the residual return from estimating a four-factor model on daily returns during 6 months prior to the starting date of the sample (October 16, 2016 to April 15, 2017). Skewness is defined as the third moment of the residual obtained from estimating a two-factor model on daily returns over the same period.

stocks with high death or jackpot probabilities have relatively high retail ownership (Conrad et al. 2014), the lottery-like proxy is consistent with the main proxy.

In the same spirit, we construct an additional proxy based on market capitalization (Barber et al. 2008; Barber and Odean 2000; Gao and Lin 2015; Kumar and Lee 2006). Market capitalization is measured as of the last trading day a month before the API shutdown (where the main sample starts). We define a dummy variable *small-cap*, which is equal to one for stocks with below-median market capitalization. Table 1 Panel A presents the correlations among the three proxies and the three sub-indices of lottery-like stocks. As they all intend to capture stocks that are preferred by active retail investors, they all positively correlate with each other.

3.2 Sample and Data

To study the impact of convenient access to financial data through APIs, we focus on retail trades made on publicly traded firms that have data available in Compustat and CRSP both before and after the shutdown of Yahoo! Finance API (May 16, 2017). Our main sample period ranges from April 16, 2017 to June 15, 2017 (inclusive), a two-month window centered around the shutdown. We also use longer and shorter windows (four months or two weeks around the shutdown) for supplemental analyses or a two-month window in 2016 (one year before the main sample period) for a falsification test.

We identify retail trades from the Trade and Quote (TAQ) database following Boehmer et al. (2021). Several recent accounting and finance studies have followed their method (Blankespoor et al. 2018; Bonsall IV et al. 2020; Bushee et al. 2020; Huang et al. 2021). The idea behind the classification is that retail trades are often executed off-exchange and offered a small price discount relative to the national best bids and offers (Boehmer et al. 2021). Specifically, we classify retail sale (buy) trades as those with TAQ exchange code "D" (indicating off-exchange trades) and prices 0.1-0.4 cents above (below) a round penny. To be more conservative, trades with prices at a round penny or near the half-penny (0.4-0.6 cents, inclusive) are not classified. Our classification misses out on some retail trades as not every retail trade is off-exchange or receives a price discount. Boehmer et al. (2021) compare this classification to a proprietary dataset of retail trades and show the retail trades identified by this classification are representative. Moreover, the measurement errors should not systematically affect our results as this classification results in similar errors both before and after the API shutdown. To make the trading volume more comparable across firms, we scale the retail trades (both buy and sell trades) by total shares outstanding and remove its normal level (the corresponding median scaled trading volume for the same day of the week over the past ten weeks) to construct the abnormal retail trading volume (*Ab_Retail_Vol*).

Table 1 Panel B splits the sample stocks by each of the treatment indicators and reports the change in retail trading volume around the API shutdown. *Pre* and *Post* indicate one month before and after the shutdown, respectively. The univariate analysis shows that both the raw and abnormal retail trades fall significantly in the post-period for retail-favored stocks. Specifically, *Retail_Vol* declines from 0.113 to 0.102 in retail-favor stocks as classified as those with above-median retail holdings. This amounts to a 9.7% (=(0.102-0.113)/0.113) decrease in a month, statistically significant at 1%. Meanwhile, retail trading for other stocks stays the same during the same period. Therefore, the univariate difference-in-differences estimate (DID, reported in the last column) is significantly negative. We observe similar significant drops in *Retail_Vol* for retail-favored stocks identified by other proxies, although the magnitude ranges from 8.6% to 10.5%. Overall, the univariate comparison depicts a consistent picture that retail trading goes down significantly in retail-favored stocks after the API shutdown.

To illustrate the changes in retail trades around the shutdown of Yahoo! Finance API, we plot the average daily retail trading volume scaled by shares outstanding (*Retail_Vol*) in Figure 1. The solid line represents retail-favored stocks and the dashed line represents other stocks. Panel A classifies stocks with above-median retail holdings as retail-favored stocks while Panels B and C use lottery-like and small-cap stocks to classify retail-favored stocks. Across the three different classifications of retail-favored stocks, we observe a similar pattern (Figure F1 exhibits similar plots for independent indices of lottery-like stocks). For starters, the solid line is always above the dashed line, confirming our identification assumption that retail trading is more active among retail-favored stocks. Despite the level differences, the two lines follow similar trends in the month before the shutdown, supporting the parallel trend assumption for DID analyses. Within three days after the shutdown, the solid line drops substantially, while the dashed line does not change noticeably except for a minor spike. The sharp decrease in retail trades immediately after the shutdown suggests that the drop in retail trades is more likely due to the shutdown than other confounding events.[10]

We obtain firm characteristics, stock performance, analyst following, and retail (institutional) ownership from standard data sources (Compustat, CRSP, IBES, and Thomas Reuters). To avoid looking-ahead biases, firm characteristics are measured at the latest fiscal year ending before January 1, 2017, hence they should be available when trading decisions are made. We use firm characteristics as reported in annual reports (4th quarter) rather than quarterly reports to increase comparability across firms as annual reports are audited, and firms may display seasonality in different quarters.

Panel A of Table 2 reports the summary statistics for the main sample, which includes about 170,000 daily observations from April 16, 2017, to June 15, 2017, for 4,209 unique firms (stocks). To minimize the influence of outliers, all variables are winsorized at 1% and 99%, except dummy variables and log-transformed variables. On a typical day, the retail trades were identified from TAQ account for 0.08% of shares outstanding. After removing the normal level of retail trading volume, the average abnormal retail trades account for 0.02% of shares outstanding. Panel B reports the firm characteristics. The average firm is modestly large and leveraged, regularly covered by media and financial analysts. The average ROA is negative, but the median ROA is slightly positive.

To measure the quality of retail trades, we study the predictive power of retail trades for future returns following the prior literature (Barber et al. 2009; Gao and Lin 2015; Kelley and Tetlock 2013; Kelley and Tetlock 2016). The idea is that buy trades (sell trades) that positively (negatively) correlate with future stock returns are likely to be more profitable on average, which reflects better investment decision-making on average. Specifically, we construct cumulative future abnormal returns (daily stock returns relative to daily market returns) over various horizons (ranging from the next week to the next two months) based on

[10] From Factiva, we downloaded 512 news articles published by the Wall Street Journal during the week of the API shutdown. After reading through the title and lead paragraphes, we did not find any significant events or incidents that would significantly affect retail investors.

CRSP data and regress the cumulative future abnormal returns on abnormal retail buy and sell trading volume. The predictive power is measured as the coefficient on the abnormal retail trading variables.

4. Empirical Analyses

4.1 Retail Trading Volume

4.1.1 Specification

We formally test the impact of the shutdown of Yahoo! Finance API on retail trading volumes using the following difference-in-differences regression specification.

$$Ab\_Retail\_Vol_{it} = \alpha + \beta \cdot Post_t \times RFS_i + \gamma \cdot W_{it} + Date\ FE + Firm\ FE + \epsilon_{it} \qquad (1)$$

where i represents the firm and t the date. The outcome variable is abnormal retail trades (*Ab_Retail_Vol*). *Post* is a dummy variable indicating the period after the shutdown of Yahoo! Finance API. The key variable of interest is the interaction term *Post* × *RFS*, whose coefficient is a DID estimate uncovering the impact of API shutdown.[11] W_{it} represents a set of firm-day level control variables including the stock return, the square of stock return, and news coverage (Da et al. 2011). Date fixed effects control for any changes in macroeconomic conditions that affect retail-favored and other stocks. Firm fixed effects control for time-invariant firm heterogeneities. In robustness checks, we also use lottery-like stocks and stocks with small market capitalization to identify the treatment group, i.e., replace *RFS* with *Lottery_Like*, *Small_Cap*.

4.1.2 Empirical Results on Retail Trading Volume

We report the regression results of abnormal retail trading volumes in Table 3. Panel A uses retail investors' revealed preferences (retail holdings) while Panel B uses lottery-like and small-cap stocks to classify retail-favored stocks. Columns 1 and 2 of Panel A use a two-week window centered around the shutdown of Yahoo! Finance API while Columns 3 and 4 (5 and 6) use a two-month (four-month) window centered around the shutdown. In Columns 1, 3, and 5, we include a set of common firm characteristics and industry fixed effects. In Columns 2, 4, and 6, firm fixed effects are included and firm characteristics are

[11] *RFS* and *Post* are not included in the regressions independently because their direct impacts are absorbed by firm and date fixed effects, respectively.

dropped as they do not change during our sample period. Consistent with the graphic evidence in Figure 1, Columns 1-4 report a negative coefficient on the interaction term (*Post*×*RFS*), significant at 1% level. This result suggests that abnormal retail trades drop significantly after the shutdown for retail-favored stocks, consistent with the idea that API induces active retail investors to trade more frequently. Once we extend the sample period to four months in Columns 5-6, the coefficient on *Post*×*RFS* becomes smaller and statistically insignificant, suggesting that active retail investors gradually find alternative data sources to substitute Yahoo! Finance API.

For control variables, we find that retail trading increases on days with larger stock movements and news coverage, consistent with the intuition that retail investors are drawn to attention-grabbing stocks (Barber and Odean 2007). Interestingly, retail investors are less active in loss firms or firms with intensive R&D, possibly because these firms are too complicated for retail investors to understand. Moreover, retail investors are more active in firms with intensive advertising expenditures, which is consistent with advertising campaigns increasing firms' visibility among retail investors (Frieder and Subrahmanyam 2005; Lou 2014).

To show the dynamic effects, we plot the weekly average DID coefficients. Specifically, we modify Model (1) by replacing the *Post* dummy with a series of dummy variables indicating each week relative to the API shutdown during the two-month sample period centered around the shutdown: $Ab\_Retail\_Vol_{it} = \alpha + \sum_{j=-4}^{4} \beta_j \cdot Week[j]_t \times RFS_i + \gamma \cdot W_{it} + Date\ FE + Firm\ FE + \epsilon_{it}$, where $Week[j]$ is 1 if the date is during week j following the shutdown (week 0 starts from the shutdown date and ends seven days later). We then plot the coefficients β_j's and the corresponding confidence intervals in Figure 2. We find no significant difference in the retail trading volumes for the control and treatment firms before the API shutdown, which further supports the parallel trend assumption.

Panel B of Table 3 reports the results using alternative proxies of retail-favored stocks, which captures the extent to which the stock's payoff structure appears like a lottery and those stocks with smaller size. Specifically, Columns 1-3 use *Lottery_Like*, a composite index of the three sub-components (*Low_Priced*, *High_Volatility*, and *High_Skewness*), as the proxy, while columns 4-6 use *small_cap*. Across the two proxies, we observe a consistent pattern that abnormal retail trading volume goes down for retail-favored stocks immediately after the API shutdown and gradually climbs back in about two months after the shutdown. Table F2 in the appendix reports similar results using separate indices of lottery-like stocks. The only exception is Column 9, where abnormal retail trading volume does not fully recover to its normal level for high-skewness stocks, although it is only marginally significant (10%).

Our DID design alleviates concerns that macroeconomic shocks drive our results as these shocks would affect retail-favored stocks and other stocks alike. The firm fixed effects control for firm-specific time-invariant heterogeneities. However, there is still plausibility that confounding events happen to some of the retail-favored stocks, and hence, all investors (active retail investors or other investors) cut trading on them. To address such concern, we examine abnormal institutional trading volume as a placebo. We follow Bushee et al. (2020) and classify institutional trades from TAQ as those non-retail trades larger than $50,000. This classification is consistent with the convention in the finance and accounting literature that assumes larger trades are likely initiated by institutional investors. We miss some institutional trades as institutional investors nowadays often break big trades into smaller ones, but in our view, this measurement error systematically biases in the undesired direction. We scale the institution trades and subtract the normal level to construct the abnormal institutional trading volume in the same fashion as the retail trades. If our main results on retail trades were caused by other events, then we should observe a similar decrease in institutional trades. That is not what we find. In Table 4, we present the results on abnormal institutional trading volume in a two-month window centered around the API shutdown. The coefficients on the interaction terms are insignificant across all different proxies of retail-favored stocks. These non-results provide us with more confidence in attributing our findings to the shutdown of API rather than other confounding events.

One may still be concerned that our results capture spurious trends. For example, retail investors are more prevailing in firms with high retail ownership and they trade less frequently in June than May as summer vacations distract them. To rule out such possibility, we repeat the same analyses on the same set of sample firms in a two-month window centered around May 16, 2016, one year before the actual shutdown. If our findings are driven by confounding events that affect the two groups differently, we should find similar results in the falsification sample. As shown in Table 5, we do not see any significant changes in abnormal retail trading volume for retail-favored stocks after the pseudo shutdown in 2016. This non-result reassures that spurious trends do not drive our main findings.

4.3 Retail Trade Quality

So far, the analyses have shown the Yahoo! Finance API shutdown affects retail trading volume, which suggests that in the pre-shutdown period, API induces active retail investors to trade more by providing them quick and convenient access to a large amount of historical data. Two reasons can explain the trading increase. On the one hand, retail investors can put the historical data access through API into good use (e.g., use the data to compare stocks and construct a diverse portfolio to reduce risk), which in turn improves the quality of their investment decisions (Gao and Lin 2015). Consequently, they trade more often and more profitably. We label this as the *information channel*. On the other hand, using the API and playing with the data may create the illusion of knowledge for these active retail investors, boost their confidence, and reduce their risk perception, which ultimately leads them to trade more often than they should (Barber and Odean 2002; Odean 1998). This implies that the retail trades are of worse quality on average, despite that they have access to more data when making the trading decisions. We label it as the *overconfidence channel*. To sum, the information channel predicts that retail trading becomes less informative on average after API shutdown as active retail investors lose some of their information sources. In contrast, the *overconfidence channel* predicts that retail trading becomes more informative on average as overconfidence lessens with the shutdown of the API.

To differentiate these two channels, we compare the collective informativeness of retail trades around the API shutdown by examining the predictivity of the extent of retail trades (we separate buys and sells due to the inherent nature of their trading) for the subsequent cumulative abnormal return. The predictivity is estimated by the Fama-Macbeth regressions as specified below. [12]

$$CAR = \alpha + \beta_1 \cdot Ab\_Retail\_Buy + \beta_2 \cdot Ab\_Retail\_Sell + \eta \cdot Z + \xi \qquad (2)$$

where the dependent variable is future cumulative abnormal returns (CAR), that is buy-and-hold return of the individual stock minus the corresponding market return, over different horizons. Our variable of interest is *Ab_Retail_Buy (sell)*, which measures abnormal buying (selling) volume by retail investors (scaled by outstanding shares) (see Appendix E for detailed definitions). A more positive (negative) β_1 (β_2) indicates that retail buys (sells) as a whole have a higher predictivity for future returns, hence more informative on average. This is a common method in the finance literature to quantify the aggregated informativeness of a given group of investors.[13] Z represents a set of firm-day level control variables (see Table 6) following Kelley and Tetlock (2013).

Table 6 reports the regression results on the aggregated informativeness of retail buys and sells made in the two-month window around the shutdown. To establish a benchmark and to see how far ahead aggregated retail trades can predict future returns, we use the entire sample (retail-favored stocks plus other stocks for the two-months window centered around the shutdown) and (for each day) measure CAR over the horizon of the subsequent week (*CAR[1W]*), Week 2 to 4 inclusive (*CAR[2W, 4W]*), and Week 5 to 8 inclusive (*CAR[5W, 8W]*). The results across the six regressions using the entire sample (columns 1-6 in Panel A) suggest the significant results are concentrated in the first month (CAR[1W] and CAR[2W, 4W]) and the significance fades as the horizon extends to the second month (CAR[5W, 8W]). Subsample analyses

[12] Fama-Macbeth regressions first run the regression separately in each period (i.e., restimate the model T times if there are T periods in the sample) then take the average of the coefficient estimates across all periods and test it against zero (Fama and MacBeth 1973). This approach avoids looking-ahead bias as each time we estimate the coefficient we never use any information from the future in the covariance matrix.

[13] An incomplete list of papers that use this method include Boehmer et al. (2008), Goetzmann and Kumar (2008), Grinblatt, Keloharju, and Linnainmaa (2012), Hvidkjaer (2008), Gao and Huang (2020), Kaniel et al. (2008), Kelley and Tetlock (2013), Kelley and Tetlock (2016).

for control and treatment groups are therefore conducted for CAR[1W, 4W], i.e, CAR for the subsequent month[14].

Next, we run the Fama-Macbeth regressions separately for retail-favored stocks and other stocks in both the pre- and post-period and perform a "difference-in-differences" comparison of the coefficients for the four subsamples (treatment_post, treatment_pre, control_post, control_pre) using Welch's t-tests. The results are presented in Panel B (RFS is classified based on retail holdings), Panel C (RFS proxied by lottery-like stocks), and Panel D (RFS proxied by small-cap stocks). Across all different proxies for RFS, we find that retail buys become relatively more profitable (or less loss-making) in the treated firms after the API shutdown. We did not find significant changes for retail sells, which may be because selling stocks could be driven by liquidity reasons, independent from the availability of API-enabled historical data. This evidence suggests the absence of API-enabled historical data filters out lower-quality retail trades, resulting in higher profitability of the average retail trade conditional on the trade taking place, consistent with the overconfidence channel.

4.4 Supplemental Analyses

In this section, we conduct a few supplemental analyses. First, we study the change in market liquidity around the shutdown of Yahoo! Finance API. Theoretical and empirical evidence shows that uninformed trading in the stock market improves market liquidity (e.g., Grossman and Miller 1988; Kelley and Tetlock 2013). More uninformed trading makes it easier for any given investor to find a counterparty to trade with. More importantly, in a market with relatively more uninformed trades, any given investor will be less concerned with adverse selection (trading with counterparties with information advantage). Consequently, investors are more willing to trade with each other, resulting in a lower price impact per share traded and lower bid-ask spread (Greene and Smart 1999; Han et al. 2016). Based on the above intuition, if historical price data induces more uninformed retail trading, we should see better liquidity before the shutdown (i.e.,

[14] For each day in the two-month window around the API shutdown, we estimate the CAR for the subsequent month. For example, for June 15, we calculate the CAR from June 16 to July 16 - the subsequent month that follows the day (not the month after the API shutdown).

liquidity deteriorates after the API shutdown). Indeed, Table 7 shows a consistent drop in liquidity (i.e., increase in illiquidity measures, *AIM* and *Spread*) for retail-favored stocks after the shutdown across different classifications of retail-favored stocks. Economically, *AIM* (*Spread*) increases by 12.3%-17.8% (5.1%-7.9%) after the shutdown for retail-favored stocks relative to other stocks. This economically considerable deterioration in market liquidity underscores the influence of API-enabled decision-making in the functioning of the overall market.

Second, we conduct two robustness checks (reported in Appendix G). To see whether the API has different impacts on retail investors when they make buy or sell transactions, we study retail buys and sells separately in Table G1. We see both retail buy and sell trades drop after the API shutdown. To sharpen the difference between stocks favored by active retail investors and other stocks, we exclude firms with the middle 20% retail holdings (above 40th percentile and below 60th percentile) and repeat the main analyses in Table G2. We find similar results for retail trading volume and trading quality.

5. Concluding Remarks

Collectively, we leverage the facts that 1) active retail investors were disproportionately affected by the Yahoo! Finance API shutdown because they relied on feedback trading built on historical price data; 2) these investors favor certain types of stocks that differ from other investors' preferences. By examining retail investments in stocks favored by active retail investors and in other stocks, we find that the low cost to access market data induces excessive trading by these investors, and to their detriment. Despite the potential to ease informational frictions, convenient access to a large amount of raw historical price data can create an illusion of knowledge and control, which exacerbates overconfidence in main-street investors, and induces more retail trades that are less predictive of future returns. These findings converge with the theoretical predictions by Zhang and Zhang (2015), who suggest that "more aggressive feedback trading creates higher risks for uninformed traders without bringing higher benefits".

The study contributes to the literature on retail investors and technology adoptions. While prior studies show that technology-induced convenience (reduction in transaction cost) can exacerbate overconfidence and excessive trading (Barber et al. 2020; Barber and Odean 2002), our findings suggest unlocking access to more information (reduction in information acquisition cost) can also create unintended consequences for retail investors. In contrast to recent studies suggesting that processed information (accounting information and investment research) enabled by financial technologies can benefit small investors (Farrell et al. 2018; Gao and Huang 2020), we show that API-enabled raw information aggravates retail investors' behavioral flaws. As technologies level the playing field for retail investors, it is important to scrutinize not only the quality (Ammann and Schaub 2020; Clarke et al. 2020) but also the type of information that is presented to retail investors.

As more information is meant to be useful, our study points to a missing piece - retail investors' financial literacy. Since information and financial judgment (acumen) complements one another in informed decision making, the current trends in democratizing access to raw data in financial markets may have inadvertently caused some retail investors to consume the data in bulks while substituting them with their lack of financial acumen. In offering retail investors more data, it is equally if not more important to ensure they have the necessary financial knowledge (Fernandes et al. 2014) to correctly use and interpret the data. As retail investors' behavioral deficiencies are amplified by the easy access to tech-enabled financial data, they have in a way become more vulnerable when exposed to such technologies. Since retail investors typically do not have an army of financial consultants to help suppress their cognitive biases (Liu et al. 2019), they are increasingly at a disadvantage when financial technologies lure them into trading more. As encouraging investors with low financial literacy into the market may not be socially beneficial (Di Maggio and Pagano 2017), our findings suggest the SEC's call to improve retail investors' financial literacy (Stein 2018) is ever more urgent with ubiquitous adoptions of financial technologies.

References

Ammann, M., and Schaub, N. 2020. "Do Individual Investors Trade on Investment-Related Internet Postings?," *Management Science* (Forthcoming).

Apesteguia, J., Oechssler, J., and Weidenholzer, S. 2020. "Copy Trading," *Management Science* (66:12), pp. 5608-5622.

Barber, B. M., Huang, X., Odean, T., and Schwarz, C. 2020. "Attention Induced Trading and Returns: Evidence from Robinhood Users." SSRN 3715077.

Barber, B. M., Lee, Y.-T., Liu, Y.-J., and Odean, T. 2008. "Just How Much Do Individual Investors Lose by Trading?," *The Review of Financial Studies* (22:2), pp. 609-632.

Barber, B. M., and Odean, T. 2000. "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *The Journal of Finance* (55:2), pp. 773-806.

Barber, B. M., and Odean, T. 2001. "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment," *The Quarterly Journal of Economics* (116:1), pp. 261-292.

Barber, B. M., and Odean, T. 2002. "Online Investors: Do the Slow Die First?," *The Review of Financial Studies* (15:2), pp. 455-488.

Barber, B. M., and Odean, T. 2007. "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors," *The Review of Financial Studies* (21:2), pp. 785-818.

Barber, B. M., and Odean, T. 2013. "The Behavior of Individual Investors," in *Handbook of the Economics of Finance*. Elsevier, pp. 1533-1570.

Barber, B. M., Odean, T., and Zhu, N. 2009. "Do Retail Trades Move Markets?," *The Review of Financial Studies* (22:1), pp. 151-186.

Blankespoor, E., Dehaan, E., Wertz, J., and Zhu, C. 2019. "Why Do Individual Investors Disregard Accounting Information? The Roles of Information Awareness and Acquisition Costs," *Journal of Accounting Research* (57:1), pp. 53-84.

Blankespoor, E., deHaan, E., and Zhu, C. 2018. "Capital Market Effects of Media Synthesis and Dissemination: Evidence from Robo-Journalism," *Review of Accounting Studies* (23:1), pp. 1-36.

Boehmer, E., Jones, C. M., and Zhang, X. 2008. "Which Shorts Are Informed?," *The Journal of Finance* (63:2), pp. 491-527.

Boehmer, E., Jones, C. M., Zhang, X., and Zhang, X. 2021. "Tracking Retail Investor Activity." *Journal of Finance* (Forthcoming).

Bonsall IV, S. B., Green, J., and Muller III, K. A. 2020. "Market Uncertainty and the Importance of Media Coverage at Earnings Announcements," *Journal of Accounting and Economics* (69:1), p. 101264.

Bushee, B., Cedergren, M., and Michels, J. 2020. "Does the Media Help or Hurt Retail Investors During the IPO Quiet Period?," *Journal of Accounting and Economics* (69:1).

Clarke, J., Chen, H., Du, D., and Hu, Y. J. 2020. "Fake News, Investor Attention, and Market Reaction," *Information Systems Research* (Forthcoming).

Conrad, J., Kapadia, N., and Xing, Y. 2014. "Death and Jackpot: Why Do Individual Investors Hold Overpriced Stocks?," *Journal of Financial Economics* (113:3), pp. 455-475.

Da, Z., Engelberg, J., and Gao, P. 2011. "In Search of Attention," *The Journal of Finance* (66:5), pp. 1461-1499.

Da, Z., Huang, X., and Jin, L. J. 2021. "Extrapolative Beliefs in the Cross-Section: What Can We Learn from the Crowds?," *Journal of Financial Economics* (140:1), pp. 175-196.

Dhar, R., and Zhu, N. 2006. "Up Close and Personal: Investor Sophistication and the Disposition Effect," *Management Science* (52:5), pp. 726-740.

Di Maggio, M., and Pagano, M. 2017. "Financial Disclosure and Market Transparency with Costly Information Processing," *Review of Finance* (22:1), pp. 117-153.

Dorn, A. J., Dorn, D., and Sengmueller, P. 2015. "Trading as Gambling," *Management Science* (61:10), pp. 2376-2393.

Fama, E. F., and MacBeth, J. D. 1973. "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy* (81:3), pp. 607-636.

Farrell, M., Green, T. C., Jame, R., and Markov, S. 2018. "The Democratization of Investment Research: Implications for Retail Investor Profitability and Firm Liquidity." SSRN 3222841.

Fernandes, D., Lynch Jr, J. G., and Netemeyer, R. G. 2014. "Financial Literacy, Financial Education, and Downstream Financial Behaviors," *Management Science* (60:8), pp. 1861-1883.

Frieder, L., and Subrahmanyam, A. 2005. "Brand Perceptions and the Market for Common Stock," *Journal of Financial and Quantitative Analysis* (40:1), pp. 57-85.

Gao, M., and Huang, J. 2020. "Informing the Market: The Effect of Modern Information Technologies on Information Production," *The Review of Financial Studies* (33:4), pp. 1367-1411.

Gao, X., and Lin, T.-C. 2015. "Do Individual Investors Treat Trading as a Fun and Exciting Gambling Activity? Evidence from Repeated Natural Experiments," *The Review of Financial Studies* (28:7), pp. 2128-2166.

Goetzmann, W. N., and Kumar, A. 2008. "Equity Portfolio Diversification," *Review of Finance* (12:3), pp. 433-463.

Graham, J. R., Harvey, C. R., and Huang, H. 2009. "Investor Competence, Trading Frequency, and Home Bias," *Management Science* (55:7), pp. 1094-1106

Greene, J., and Smart, S. 1999. "Liquidity Provision and Noise Trading: Evidence from the 'Investment Dartboard' Column," *The Journal of Finance* (54:5), pp. 1885-1899.

Grinblatt, M., and Keloharju, M. 2009. "Sensation Seeking, Overconfidence, and Trading Activity," *The Journal of Finance* (64:2), pp. 549-578.

Grinblatt, M., Keloharju, M., and Linnainmaa, J. T. 2012. "IQ, Trading Behavior, and Performance," *Journal of Financial Economics* (104:2), pp. 339-362.

Grossman, S. J., and Miller, M. H. 1988. "Liquidity and Market Structure," *The Journal of Finance* (43:3), pp. 617-633.

Han, B., Tang, Y., and Yang, L. 2016. "Public Information and Uninformed Trading: Implications for Market Liquidity and Price Efficiency," *Journal of Economic Theory* (163), pp. 604-643.

Huang, S., O'Hara, M., and Zhong, Z. 2021. "Innovation and Informed Trading: Evidence from Industry ETFs," *The Review of Financial Studies* (34:3), pp. 1280-1316.

Hvidkjaer, S. 2008. "Small Trades and the Cross-Section of Stock Returns," *The Review of Financial Studies* (21:3), pp. 1123-1151.

Ivković, Z., Sialm, C., and Weisbenner, S. 2008. "Portfolio Concentration and the Performance of Individual Investors," *Journal of Financial and Quantitative Analysis* (43:3), pp. 613-655.

Kaniel, R., Saar, G., and Titman, S. 2008. "Individual Investor Trading and Stock Returns," *The Journal of Finance* (63:1), pp. 273-310.

Kelley, E. K., and Tetlock, P. C. 2013. "How Wise Are Crowds? Insights from Retail Orders and Stock Returns," *The Journal of Finance* (68:3), pp. 1229-1265.

Kelley, E. K., and Tetlock, P. C. 2016. "Retail Short Selling and Stock Prices," *The Review of Financial Studies* (30:3), pp. 801-834.

Kumar, A. 2009. "Who Gambles in the Stock Market?," *The Journal of Finance* (64:4), pp. 1889-1933.

Kumar, A., and Lee, C. M. 2006. "Retail Investor Sentiment and Return Comovements," *The Journal of Finance* (61:5), pp. 2451-2486.

Lawrence, A., Ryans, J., Sun, E., and Laptev, N. 2016. "Yahoo Finance Search and Earnings Announcements." SSRN 2804353.

Liu, C., Mithas, S., Hsieh, P.-A., and Pan, Y. 2019. "The Effect of Financial Advisers on Investors' Gambling Preference in the Digital Era," in: *Conference on Information Systems and Technology (CIST)*. Seattle, WA.

Lou, D. 2014. "Attracting Investor Attention through Advertising," *The Review of Financial Studies* (27:6), pp. 1797-1829.

Lusardi, A., and Mitchell, O. S. 2014. "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature* (52:1), pp. 5-44.

Newey, W. K., and West, K. D. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelationconsistent Covariance Matrix," *Econometrica* (55:3), pp. 703-708.

Odean, T. 1998. "Volume, Volatility, Price, and Profit When All Traders Are above Average," *The Journal of Finance* (53:6), pp. 1887-1934.

Sias, R. W., and Starks, L. T. 1997. "Institutions and Individuals at the Turn-of-the-Year," *The Journal of Finance* (52:4), pp. 1543-1562.

Stein, K. 2018. "Remarks at SEC Speaks: Increasing Product Complexity: What's at Stake?", from https://www.sec.gov/news/speech/stein-sec-speaks-increasing-product-complexity

Von Gaudecker, H.-M. 2015. "How Does Household Portfolio Diversification Vary with Financial Literacy and Financial Advice?," *The Journal of Finance* (70:2), pp. 489-507.

Zhang, X. M., and Zhang, L. 2015. "How Does the Internet Affect the Financial Market? An Equilibrium Model of Internet-Facilitated Feedback Trading," *MIS Quarterly* (39:1), pp. 17-37.

Figure 1. Retail trading volume around the shutdown of Yahoo! Finance API

This figure plots daily retail trading volume around the shutdown of Yahoo! Finance API for retail-favored stocks (solid line) and other stocks (dashed line), respectively. The y-axis is retail trading volume scaled by total shares outstanding, multiplied by 100. In Panel A, we use retail holding to proxy for the revealed preference of retail investors. Panels B and C designate lottery-like stocks and small-cap stocks as alternative proxies for retail-favored stocks. The vertical dashed lines indicate the shutdown of Yahoo! Finance API.

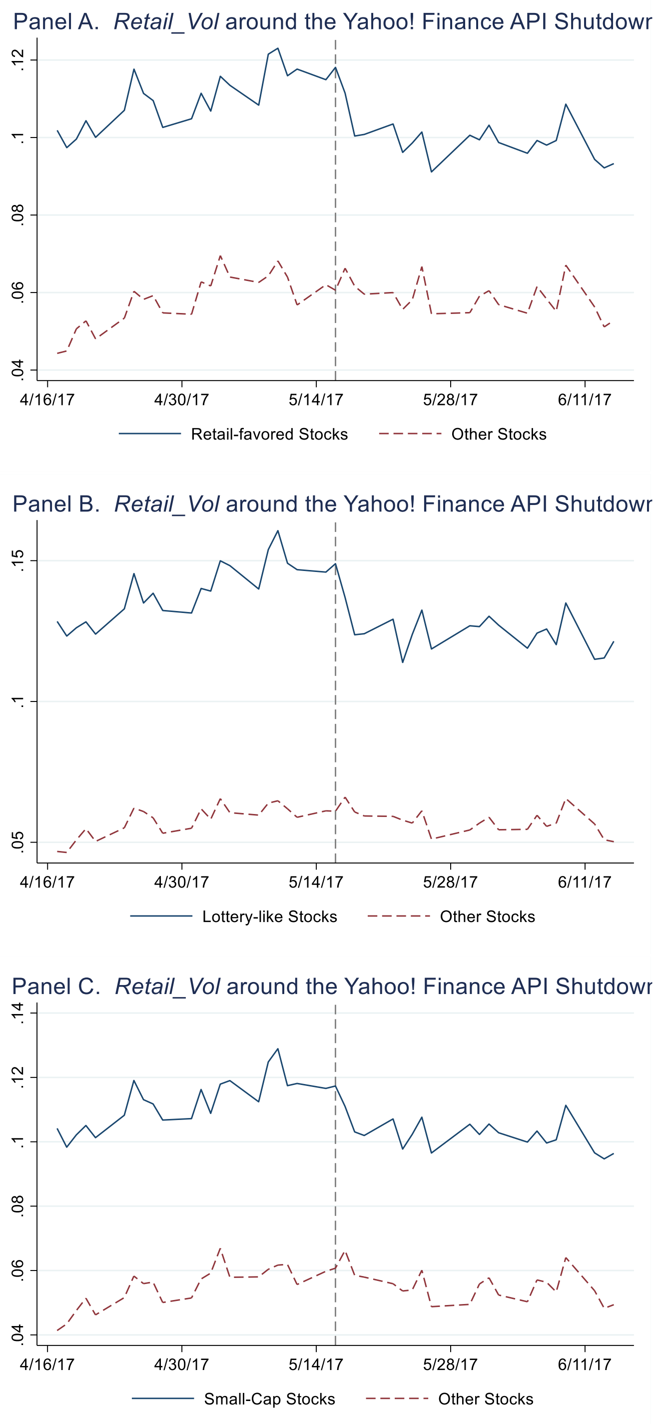


Figure 2. Weekly DID estimates

This figure plots the estimated difference in abnormal retail trading volume between retail-favored stocks and other stocks. The dots are the point estimates for RFS, and the vertical lines are the corresponding 90% confidence intervals of the weekly average DID coefficients. Specifically, we regress abnormal retail trading volume on the interaction of *RFS* and a series of dummy variables indicating each week relative to the API shutdown during the two-month sample period centered around the shutdown. The regression includes firm fixed effects, date fixed effects, and daily control variables (Return, Return^2, daily news coverage). 0W is the first week when the Yahoo! Finance API was shut down.

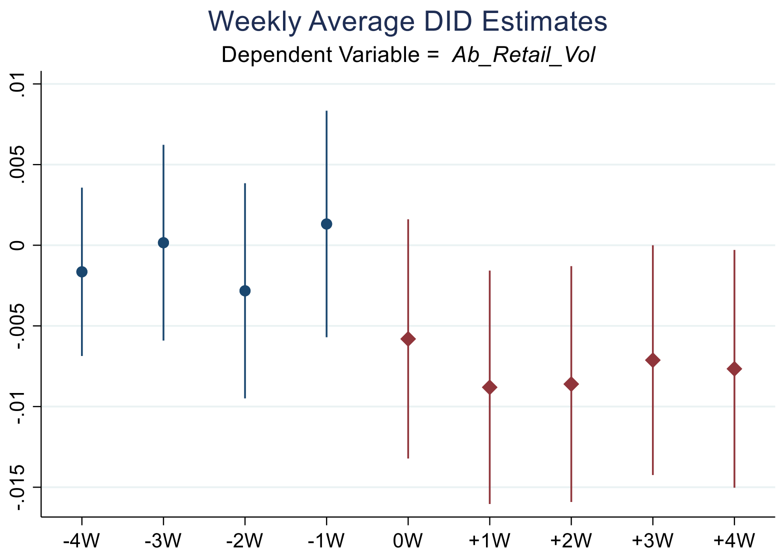


Table 1. Retail-favored, lottery-like, and small-cap stocks

Panel A of this table presents the correlations between retail-favored, lottery-like (different proxies), and small-cap stocks. Panel B presents the univariate comparisons between retail-favored/lottery-like/small-cap stocks and other stocks around the shutdown of Yahoo! Finance API. *Pre* and *Post* indicate the sub-periods before and after the shutdown, respectively. See Appendix E for detailed variable definitions. ***, **, and * stand for statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A. Correlation between different proxies to identify retail-favored stocks

	RFS	*Lottery_Like*	Low_Priced	High_Volatility	High_Skewness
Lottery_Like	0.322***				
Low_Priced	0.403***	0.634***			
High_Volatility	0.234***	0.452***	0.430***		
High_Skewness	0.144***	0.561***	0.121***	0.133***	
Small_Cap	0.504***	0.500***	0.598***	0.467***	0.201***

Panel B. Univariate Comparisons

	Pre	Post	MeanDiff	Pre	Post	MeanDiff	DID
		RFS = 1			**RFS = 0**		
Retail_Vol	0.113	0.102	-0.011***	0.059	0.059	0.000	-0.011***
Ab_Retail_Vol	0.027	0.020	-0.007***	0.015	0.015	0.000	-0.007**
		Lottery_Like = 1			**Lottery_Like = 0**		
Retail_Vol	0.143	0.129	-0.015***	0.059	0.058	0.001	-0.014***
Ab_Retail_Vol	0.033	0.024	-0.009**	0.015	0.015	0.000	-0.009**
		Small_Cap = 1			**Small_Cap = 0**		
Retail_Vol	0.116	0.105	-0.010***	0.056	0.056	0.000	-0.011***
Ab_Retail_Vol	0.029	0.022	-0.006***	0.013	0.013	0.000	-0.007**

Table 2. Summary statistics

Panel A of this table reports the summary statistics of the key variables used in the main sample of this study (a two-month window centered around the shutdown of Yahoo! Finance API). Each observation is a firm-trading day for daily measures. Panel B presents common firm characteristics measured as of the most recent fiscal year before the sample starting date. See Appendix E for detailed variable definitions.

Panel A. Firm-day obs.

	N	Mean	S.D.	P25	P50	P75
Retail_Vol	169430	0.083	0.200	0.010	0.025	0.065
Ab_Retail_Vol	169430	0.019	0.131	-0.008	0.000	0.014
AIM	168921	0.073	0.246	0.000	0.002	0.017
Spread	169428	0.509	0.990	0.037	0.117	0.456
Institutional_Vol	169430	0.081	0.168	0.000	0.020	0.083
Ab_Institutional_Vol	169430	0.032	0.134	-0.009	0.000	0.025
Ret	169430	0.000	0.023	-0.010	0.000	0.011
Ret^2	169430	0.001	0.002	0.000	0.000	0.000
News	169430	0.225	0.535	0.000	0.000	0.000

Panel B. Firm-level obs.

	N	Mean	S.D.	P25	P50	P75
ROA	4209	-0.068	0.287	-0.041	0.013	0.053
Loss	4209	0.344	0.475	0.000	0.000	1.000
R&D	4209	0.065	0.156	0.000	0.000	0.045
Advertising	4209	0.009	0.025	0.000	0.000	0.003
Leverage	4209	0.260	0.242	0.048	0.216	0.406
Analysts	4209	0.968	0.937	0.000	0.693	1.609

Table 3. Retail trades around the shutdown of Yahoo! Finance API

Panel A of this table reports the regression results of retail trades around the shutdown of Yahoo! Finance API (May 16, 2017). The sample is a panel of firm-days in a two-week, two-, or four-month window (indicated in the table header) centered around the shutdown. The dependent variable is the abnormal retail trading volume (*Ab_Retail_Vol*). The key variable of interest is the interaction between *Post* and retail-favored stocks *(RFS)*. In Panel B, we use lottery-like stocks (*Lottery_like*, a dummy variable equal to one for stocks with low price, high volatility, and high skewness) and small-cap stocks (*Small_Cap*, a dummy variable equal to one for stocks with below-median market capitalization) as alternative proxies for RFS. See Appendix E for detailed variable definitions. Robust standard errors clustered by firm are reported in the parentheses. ***, **, and * stand for statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A. Retail-Preferred stocks measured by holding by retail investors

	(1)	(2)	(3)	(4)	(5)	(6)
DV = *Ab_Retail_Vol*	2-Week Window		2-Month Window		4-Month Window	
Post×RFS	-0.008***	-0.008***	-0.007***	-0.007***	-0.002	-0.002
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Ret	0.154***	0.165***	0.111***	0.134***	0.141***	0.151***
	(0.057)	(0.045)	(0.030)	(0.027)	(0.022)	(0.020)
Ret^2	37.594***	30.449***	36.385***	33.510***	37.215***	35.815***
	(1.402)	(1.085)	(0.995)	(0.820)	(0.930)	(0.821)
News	0.021***	0.021***	0.018***	0.022***	0.018***	0.023***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
RFS	0.003		0.001		-0.002	
	(0.003)		(0.002)		(0.002)	
Size	-0.002		-0.001		-0.000	
	(0.001)		(0.001)		(0.001)	
BTW	-0.001		-0.000		0.002	
	(0.003)		(0.002)		(0.001)	
ROA	0.007		-0.005		-0.015*	
	(0.013)		(0.009)		(0.008)	
Loss	-0.001		-0.004*		-0.006***	
	(0.004)		(0.003)		(0.002)	
R&D	-0.034*		-0.026*		-0.017	
	(0.020)		(0.015)		(0.012)	
Advertising	0.314***		0.194***		0.102***	
	(0.094)		(0.057)		(0.037)	
Leverage	0.006		0.001		-0.001	
	(0.007)		(0.004)		(0.003)	
Analysts	0.002		0.002		0.001	
	(0.002)		(0.001)		(0.001)	
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	No	Yes	No	Yes	No
Firm FE	No	Yes	No	Yes	No	Yes
Observations	43,374	43,374	169,430	169,430	326,675	326,675
R-squared	0.230	0.542	0.211	0.362	0.213	0.291

Panel B. Alternative proxies for retail-favored stocks

DV = *Ab_Retail_Vol*	(1) 2-Week	(2) 2-Month	(3) 4-Month	(4) 2-Week	(5) 2-Month	(6) 4-Month
Post× Lottery_Like	-0.014***	-0.009***	-0.003			
	(0.004)	(0.003)	(0.003)			
Post× Small_Cap				-0.009***	-0.006***	-0.003
				(0.003)	(0.002)	(0.002)
Ret	0.158***	0.134***	0.151***	0.159***	0.134***	0.151***
	(0.046)	(0.027)	(0.020)	(0.046)	(0.027)	(0.020)
Ret^2	29.812***	33.511***	35.815***	29.808***	33.506***	35.820***
	(1.094)	(0.820)	(0.821)	(1.094)	(0.819)	(0.821)
News	0.021***	0.022***	0.023***	0.021***	0.022***	0.023***
	(0.002)	(0.001)	(0.001)	-0.009***	-0.006***	-0.003
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	39,433	169,426	326,635	39,433	169,426	326,627
R-squared	0.558	0.362	0.291	0.557	0.362	0.291

Table 4. Placebo tests using institutional trades

This table reports the regression results of institutional trades around the shutdown of Yahoo! Finance API (May 16, 2017). The sample is a panel of firm-days in a two-month window centered around the shutdown. The dependent variable is the abnormal institutional trading volume (*Ab_Institutional_Vol*). The regression specification is the same as in Table 3. The key variable of interest is the interaction between *Post* (indicating the period after the shutdown of Yahoo! Finance API) and *Treat* (stocks preferred by retail investors based on their holdings (*RFS*), lottery-like stocks (*Lottery_like*), or small-cap stocks (*Small_Cap*), as indicated in the table header). See Appendix E for detailed variable definitions. Robust standard errors clustered by firm are reported in the parentheses. ***, **, and * stand for statistical significance at the 1%, 5%, and 10% level, respectively.

DV = *Ab_ Institutional _Vol*	(1)	(2)	(3)
Treat =	RFS	Lottery_Like	Samll_Cap
Post×Treat	0.002	0.000	0.002
	(0.002)	(0.002)	(0.002)
Ret	0.040*	0.041**	0.041*
	(0.021)	(0.021)	(0.021)
Ret^2	18.582***	18.849***	18.722***
	(0.618)	(0.626)	(0.622)
News	0.037***	0.037***	0.037***
	(0.001)	(0.001)	(0.001)
Day FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	169,430	169,211	169,426
R-squared	0.194	0.194	0.194

Table 5. Falsification tests

This table reports the results of falsification tests in a two-month window centered around May 16, 2016, one year before the shutdown of Yahoo! Finance API. The dependent variable is the abnormal retail trading volume (*Ab_Retail_Vol*). The regression specification is the same as in Table 3. The key variable of interest is the interaction between *Post* (indicating the period after the shutdown of Yahoo! Finance API) and *Treat* (stocks preferred by retail investors based on their holdings (*RFS*), lottery-like stocks (*Lottery_like*), or small-cap stocks (*Small_Cap*), as indicated in the table header). See Appendix E for detailed variable definitions. Robust standard errors clustered by firm are reported in the parentheses. ***, **, and * stand for statistical significance at the 1%, 5%, and 10% level, respectively.

DV = *Ab_ Retail _Vol*	(1)	(2)	(3)
Treat =	RFS	Lottery_Like	Small_Cap
Post×Treat	0.002	-0.001	0.001
	(0.002)	(0.002)	(0.002)
Ret	0.027	0.027*	0.025
	(0.016)	(0.016)	(0.016)
Ret^2	21.346***	21.347***	21.372***
	(0.506)	(0.506)	(0.507)
News	0.020***	0.020***	0.020***
	(0.001)	(0.001)	(0.001)
Day FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	156,593	156,593	156,467
R-squared	0.362	0.362	0.362

Table 6. The predictivity of retail trades for future returns

This table reports the daily Fama-Macbeth regressions of future returns on abnormal retail buy and sell trading volume. The sample includes firm-day observations during a 2-month window centered around the Yahoo! Finance API shutdown. The dependent variable of Panel A is the buy-and-hold abnormal returns for the next week starting from the next day in Columns 1-2, from week 2 to week 4 in Columns 3-4, and from week 5 to week 8 in Columns 5-6. The key variable of interest is abnormal retail buys and sells (*Ab_Retail_Buy*, *Ab_Retail_Sell*). In Panel B, we run separate regressions in retail-favored stocks (*RFS*) and other stocks. In Panel C (D), we run separate analyses in lottery-like (small-cap) stocks and other stocks. See Appendix E for detailed variable definitions. Newey and West (1987) standard errors with lags of two are reported in the parentheses. ***, **, and * stand for statistical significance at the 1%, 5%, and 10%, respectively.

Panel A. The predictivity of retail trades for future returns

	(1)	(2)	(3)	(4)	(5)	(6)
DV =	CAR[1W] &		CAR[2W,4W] &		CAR[5W,8W] &	
Sample	All		All		All	
	Pre	Post	Pre	Post	Pre	Post
Ab_Retail_Buy	0.302	1.582**	-0.498	2.869*	-0.278	1.930
	(0.704)	(0.656)	(1.451)	(1.519)	(1.756)	(1.496)
Ab_Retail_Sell	-2.272***	-2.943***	-1.251	-2.443**	-3.310	2.200
	(0.680)	(0.692)	(1.282)	(1.085)	(2.728)	(1.399)
Ret[0]^2	-109.234***	-120.145***	-96.287**	-25.049	-62.303	-214.642**
	(29.723)	(28.015)	(39.199)	(56.225)	(90.675)	(76.750)
News	0.067	0.061	0.123	-0.011	0.199**	0.021
	(0.051)	(0.095)	(0.104)	(0.056)	(0.072)	(0.065)
Size	0.070*	0.063	0.284***	-0.053	0.027	0.364**
	(0.035)	(0.042)	(0.036)	(0.082)	(0.082)	(0.138)
BTM	-0.078	0.023	0.087	0.074	0.177	0.567***
	(0.114)	(0.148)	(0.099)	(0.089)	(0.152)	(0.183)
Ret[0]	3.969	-6.458**	3.377	-5.100	11.960*	0.726
	(2.856)	(2.765)	(2.926)	(4.407)	(6.300)	(2.916)
CAR[-1W]	0.018	-0.017	0.057**	-0.034***	0.052	0.020
	(0.016)	(0.014)	(0.024)	(0.012)	(0.049)	(0.016)
CAR[-2W,-4W]	0.005	0.009	0.085***	-0.007	-0.018	0.047***
	(0.009)	(0.015)	(0.009)	(0.010)	(0.013)	(0.009)
Observations	82,786	86,350	82,786	86,350	82,654	86,102
R-squared	0.017	0.026	0.022	0.014	0.018	0.024

Panel B. The return predictivity of retail trades in RFS and other stocks

	(1)	(2)	(3)	(4)
DV =	CAR[1W,4W]		CAR[1W,4W]	
Sample	RFS = 0		RFS = 1	
	Pre	Post	Pre	Post
Ab_Retail_Buy	7.566**	-0.943	-1.022	7.303***
	(3.512)	(2.542)	(1.622)	(1.383)
Ab_Retail_Sell	-2.365	-3.985	-5.176***	-6.349***
	(3.285)	(3.824)	(1.146)	(1.194)
Ab_Retail_Buy (Col 4-Col 3) – (Col 2-Col 1):		16.834***		
Ab_Retail_Sell (Col 4-Col 3) – (Col 2-Col 1):		0.447		
Controls	Yes	Yes	Yes	Yes
Observations	41,837	43,645	40,949	42,705
R-squared	0.032	0.033	0.020	0.015

Panel C. The return predictivity of retail trades in lottery-like and other stocks

	(1)	(2)	(3)	(4)
DV =	CAR[1W,4W]		CAR[1W,4W]	
Sample	Lottery_Like = 0		Lottery_Like = 1	
	Pre	Post	Pre	Post
Ab_Retail_Buy	5.081*	1.366	-1.457	7.768***
	(2.918)	(2.426)	(2.020)	(1.144)
Ab_Retail_Sell	-4.613**	-8.167***	-4.308***	-4.823***
	(2.187)	(1.910)	(1.232)	(1.502)
Ab_Retail_Buy (Col 4-Col 3) – (Col 2-Col 1):		12.94***		
Ab_Retail_Sell (Col 4-Col 3) – (Col 2-Col 1):		3.039		
Controls	Yes	Yes	Yes	Yes
Observations	56,418	58,857	26,368	27,489
R-squared	0.020	0.031	0.024	0.017

Panel D. The return predictivity of retail trades in small-cap and other stocks

DV =	CAR[1W,4W]		CAR[1W,4W]	
Sample	Small_Cap = 0		Small_Cap = 1	
	Pre	Post	Pre	Post
Ab_Retail_Buy	5.216	2.399	-0.121	6.730***
	(3.226)	(2.570)	(1.850)	(1.341)
Ab_Retail_Sell	-6.897**	-6.548*	-4.186***	-6.156***
	(2.908)	(3.301)	(1.377)	(1.214)
Ab_Retail_Buy (Col 4-Col 3) – (Col 2-Col 1):		9.668**		
Ab_Retail_Sell (Col 4-Col 3) – (Col 2-Col 1):		-2.319		
Controls	Yes	Yes	Yes	Yes
Observations	41,387	43,231	41,399	43,115
R-squared	0.038	0.047	0.016	0.020

Table 7. Market liquidity

This table reports the regression results of market liquidity in a two-month window centered around the shutdown of Yahoo! Finance API. The dependent variable is the daily Amihud's illiquidity measure (*AIM*) in Panel A and the daily relative bid-ask spread (*Spread*) in Panel B. The key variable of interest is the interaction between *Post* (indicating the period after the shutdown of Yahoo! Finance API) and *Treat* (stocks preferred by retail investors based on their holdings (*RFS*), lottery-like stocks (*Lottery_like*), or small-cap stocks (*Small_Cap*), as indicated in the table header). See Appendix E for detailed variable definitions. Robust standard errors clustered by firm are reported in the parentheses. ***, **, and * stand for statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A. AIM

DV = *AIM*	(1)	(2)	(3)
Treat =	RFS	Lottery_Like	Small_Cap
Post×Treat	0.009***	0.013***	0.009***
	(0.002)	(0.003)	(0.002)
Ret	-0.218***	-0.217***	-0.218***
	(0.025)	(0.025)	(0.025)
Ret^2	16.130***	16.130***	16.137***
	(0.727)	(0.727)	(0.727)
News	-0.012***	-0.012***	-0.012***
	(0.001)	(0.001)	(0.001)
Day FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	168,921	168,917	168,917
R-squared	0.612	0.612	0.612

Panel B. Spread

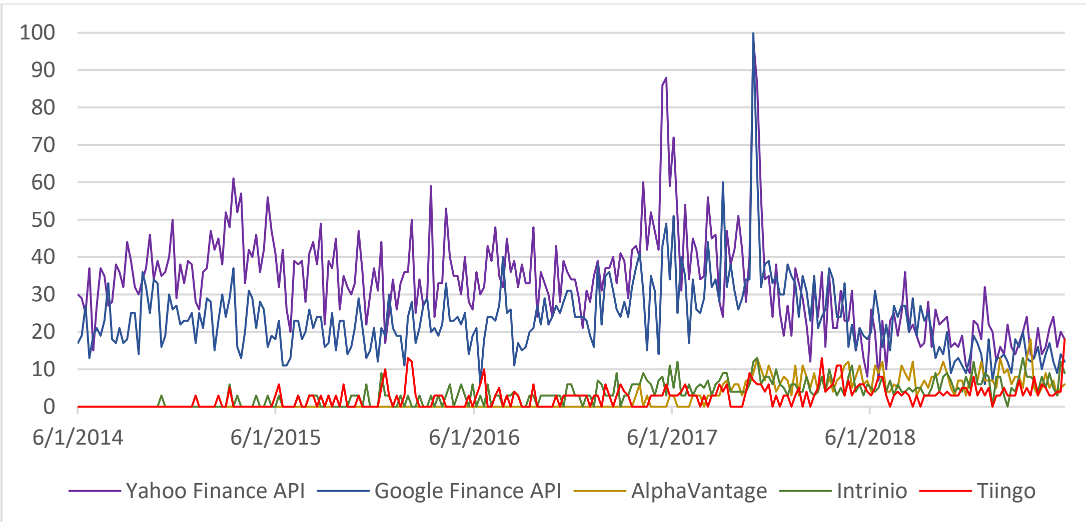
DV = *Spread*	(1)	(2)	(3)
Treat =	RFS	Lottery_Like	Small_Cap
Post×Treat	0.026***	0.040***	0.032***
	(0.009)	(0.012)	(0.009)
Ret	0.146	0.149*	0.148*
	(0.089)	(0.089)	(0.089)
Ret^2	13.103***	13.103***	13.128***
	(1.774)	(1.773)	(1.775)
News	-0.009***	-0.010***	-0.010***
	(0.002)	(0.002)	(0.002)
Day FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	169,428	169,424	169,424
R-squared	0.753	0.753	0.753

Appendix A. Yahoo! Finance API and alternative APIs

While the data on the exact market share of Yahoo! Finance API is unavailable, we use Google search volumes to estimate its popularity. Based on various forum posts (see Appendix C), we identified four main alternatives of Yahoo! Finance API, namely Google Finance API, AlphaVantage, Intrinio, and Tiingo. The Google trend index measures the relative search frequencies of one or more keywords in a given period. Figure A1 shows that Yahoo! Finance API was consistently the top financial API until its shutdown in mid-2017.

Figure A1. Worldwide Google Search Volume Index on Finance APIs

This figure depicts the worldwide weekly Google search volume index from 2014 to 2019 for the five popular Finance APIs. The highest search volume in the period is assigned the score of 100. The search volume for Yahoo! Finance API is represented by the purple line.



Appendix B. Financial variables accessible through Yahoo! Finance API

Yahoo! Finance provides historical and real-time stock quotes in CSV files. The API allows users to access those data in large bulks that are otherwise impractical to do so manually. Based on a tutorial for Yahoo Finance API[15], Table B1 exhibits a non-exhaustive list of financial variables accessible through the API.

Table B1. Sample variables accessible through Yahoo! Finance API

Pricing	**Dividends**
a: Ask	**y**: Dividend Yield
b: Bid	**d**: Dividend per Share
b2: Ask (Realtime)	**r1**: Dividend Pay Date
b3: Bid (Realtime)	**q**: Ex-Dividend Date
p: Previous Close	
o: Open	
Date	
c1: Change	**d1**: Last Trade Date
c: Change & Percent Change	**d2**: Trade Date
c6: Change (Realtime)	**t1**: Last Trade Time
k2: Change Percent (Realtime)	
p2: Change in Percent	
Averages	
c8: After Hours Change (Realtime)	**m5**: Change From 200 Day Moving Average
c3: Commission	**m6**: Percent Change From 200 Day Moving Average
g: Day's Low	**m7**: Change From 50 Day Moving Average
h: Day's High	**m8**: Percent Change From 50 Day Moving Average
k1: Last Trade (Realtime) With Time	**m3**: 50 Day Moving Average
l: Last Trade (With Time)	**m4**: 200 Day Moving Average
l1: Last Trade (Price Only)	
t8: 1 yr Target Price	
Misc	
w1: Day's Value Change	**g1**: Holdings Gain Percent
w4: Day's Value Change (Realtime)	**g3**: Annualized Gain
p1: Price Paid	**g4**: Holdings Gain
m: Day's Range	**g5**: Holdings Gain Percent (Realtime)
m2: Day's Range (Realtime)	**g6**: Holdings Gain (Realtime)
52 Week Pricing	**Symbol Info**
k: 52 Week High	**i**: More Info
j: 52 week Low	**j1**: Market Capitalization
j5: Change From 52 Week Low	**j3**: Market Cap (Realtime)
k4: Change From 52 week High	**f6**: Float Shares
j6: Percent Change From 52 week Low	**n**: Name
k5: Percent Change From 52 week High	**n4**: Notes
w: 52 week Range	**s**: Symbol
	s1: Shares Owned
	x: Stock Exchange
	j2: Shares Outstanding
Volume	
v: Volume	
a5: Ask Size	
b6: Bid Size	**Misc**
k3: Last Trade Size	**t7**: Ticker Trend
a2: Average Daily Volume	**t6**: Trade Links

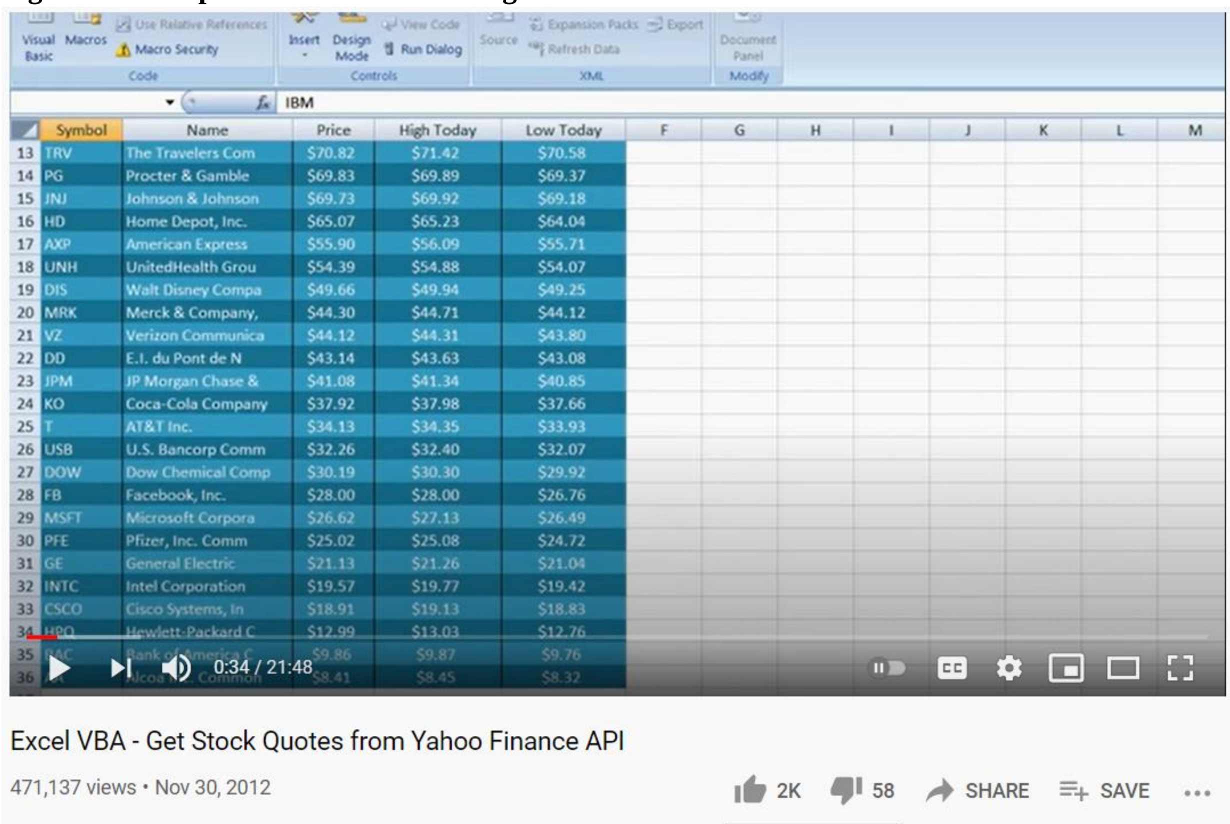
[15] https://web.archive.org/web/20171021180558/http:/www.jarloo.com/yahoo_finance/

	i5: Order Book (Realtime)
Ratios	**l2**: High Limit
e: Earnings per Share	**l3**: Low Limit
e7: EPS Estimate Current Year	**v1**: Holdings Value
e8: EPS Estimate Next Year	**v7**: Holdings Value (Realtime)
e9: EPS Estimate Next Quarter	**s6** Revenue
b4: Book Value	
j4: EBITDA	
p5: Price / Sales	
p6: Price / Book	
r: P/E Ratio	
r2: P/E Ratio (Realtime)	
r5: PEG Ratio	
r6: Price / EPS Estimate Current Year	
r7: Price / EPS Estimate Next Year	
s7: Short Ratio	

Appendix C. Userbase of Yahoo! Finance API

Yahoo! Finance API has a sizable userbase as evidenced by the number of related YouTube tutorial videos, the number of views of questions related to the API on Stack Overflow and Quora, as well as discussions on Yahoo! Help Community. The screenshots were captured in March 2021.

Figure C1. Sample Videos on Connecting to Yahoo! Finance API



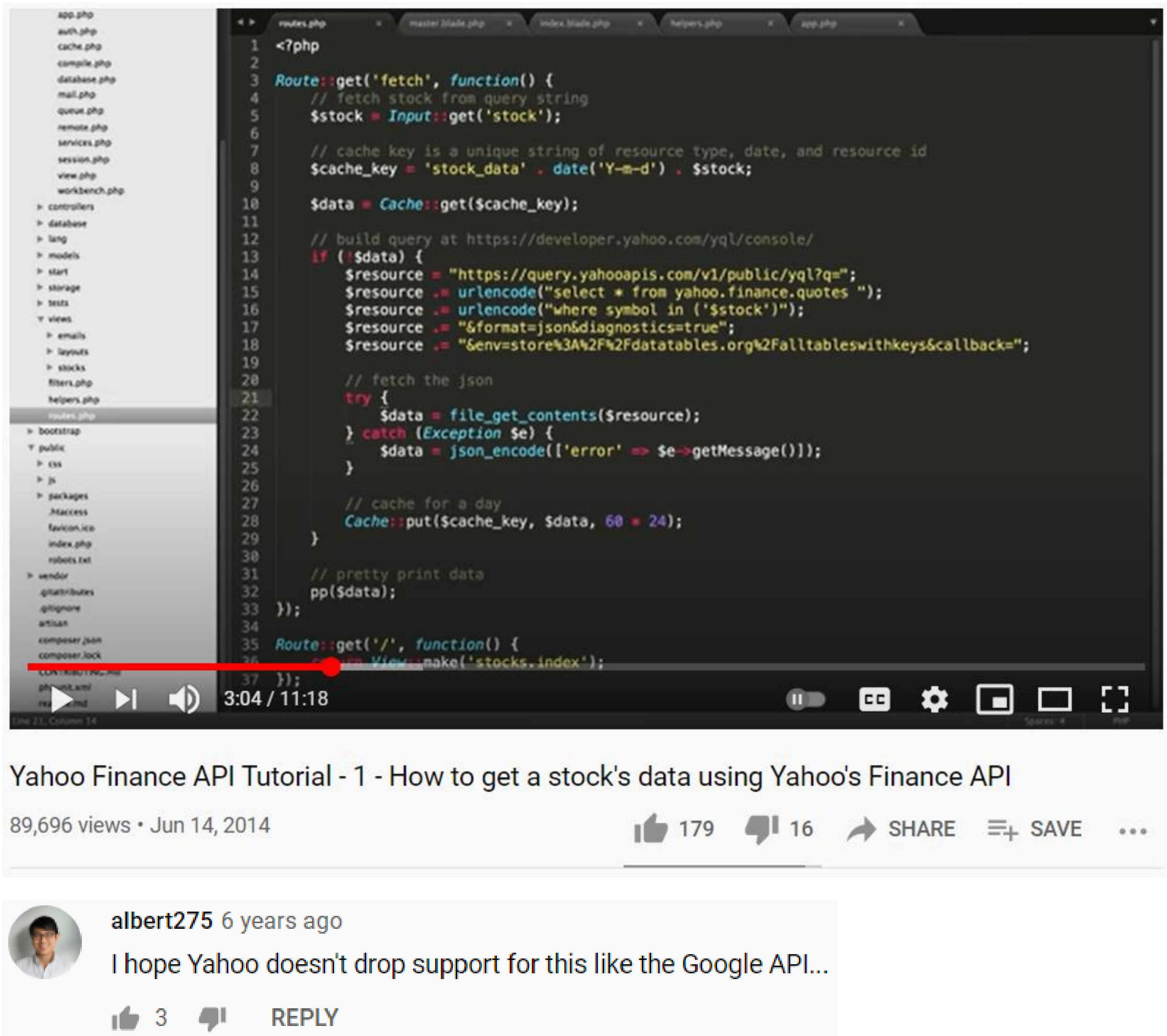


Table C1. Viewership for YouTube videos related to Finance APIs

This table reports the total number of views, comments, likes, and dislikes for the top ten YouTube videos (sorted by relevance) related to (Yahoo!) Finance APIs as of March 2021. The upload dates of the videos span from Dec 1, 2012 to May 3, 2017.

Keyword	Total Views	Total Comments	Total Likes	Total Dislikes
Finance API	4,375,348	2,307	62,772	1,174
Yahoo Finance API	1,386,800	1,293	11,194	202

Figure C2. Sample Q&As about Yahoo! Finance API on Stack Overflow

Yahoo Finance API [closed]

Asked 11 years, 3 months ago Active 5 years, 11 months ago Viewed 102k times

41
votes

48

Closed. This question does not meet Stack Overflow guidelines. It is not currently accepting answers.

Want to improve this question? Update the question so it's on-topic for Stack Overflow.
Closed 5 years ago.

Locked. This question and its answers are locked because the question is off-topic but has historical significance. It is not currently accepting new answers or interactions.

Q. Do Yahoo provides any Finance API? If yes, the what's the link to that API.

java yahoo-finance

Share

edited Jun 10 '14 at 19:36
Kara
5,609 • 15 • 47 • 55

asked Nov 19 '09 at 13:30
Yatendra
31k • 88 • 208 • 290

Also, consider the Mergent financial data API: mergent.com/servius – Eugene Osovetsky Aug 2 '10 at 20:25

You can get CSV files for up to 200 symbols using there free API jarloo.com/yahoo_finance – Kelly Feb 25 '12 at 20:21

4 @DaveWebb Google's Finance API has been deprecated and will be gone by October 2012 :(– jm3 May 12 '12 at 14:55

11 Google Finance API is deprecated now ... :(– bouncingHippo Nov 9 '12 at 19:51

For whatever it's worth, I try to keep an updated list of Finance API's here. It would seem Yahoo and Google both have API's available as of Aug 26, 2013, though I don't know licensing on them. – josephdpurcell Aug 26 '13 at 21:51

Comments disabled on deleted / locked posts / reviews

Table C2. Viewership for Popular Stack Exchange Questions related to Yahoo! Finance APIs

This table reports the view count and answer count of popular Stack Exchange questions about Yahoo! Finance APIs (as of March 2021). The questions are ordered chronologically. We observe a spike in the number of questions and views around the shutdown of the API in 2017. To put the numbers into perspective, the average view per question on Stack Overflow is fewer than 2000[16].

Question Date	Question Title	View Count	Answer Count
11/19/2009	Yahoo Finance API	102,000	4
9/26/2012	yahoo finance api returns empty response	2,853	1
1/5/2014	Retrieving Stock Quotes using Yahoo Finance API	5,755	2
3/12/2014	Yahoo! Finance API, how to get historical intraday data for one particular day?	20,145	2
4/17/2015	Yahoo Finance API all stocks?	2,097	2
10/2/2015	Yahoo Finance API stock/ticker lookup only allowing exact match	8,027	3
10/20/2015	Using Yahoo Finance API for Minute Data?	5,117	2
11/8/2015	How to pull "Last Trade Date" data from Yahoo! Finance API within Excel?	998	1
2/13/2016	Live currency rate using Yahoo Finance API	1,198	4
7/14/2016	Yahoo Finance API for BSE & NSE	1,627	1
8/1/2016	How to specify the date in a yahoo finance api query?	1,065	0
4/20/2017	Yahoo Finance API Java Download CSV	2,933	5
4/21/2017	java Yahoo finance api not returning historical data	1,162	1
5/18/2017	Yahoo Finance API changes (2017)	87,518	3
5/22/2017	Yahoo Finance API / URL not working: Python fix for Pandas DataReader	12,809	4
5/27/2017	How to use getReturns with the Yahoo finance API	5,784	3
9/14/2017	Yahoo Finance API get list of all mutual funds and ETFs tickers	1,493	1
9/19/2017	Yahoo finance API and excel vba	4,529	4
11/27/2017	Alternative to yahoo finance api	7,689	4
3/3/2018	Pandas DataReader is no longer working with the Yahoo Finance API?	1,534	2

[16] https://meta.stackexchange.com/questions/284139/what-is-the-average-number-of-views-per-question

Figure C3. Sample Q&As about Yahoo! Finance API on Quora

Did Yahoo Stock API shut down?

Answer · Follow · 25 · Request

14 Answers

Denis Alaev, more than 7 years in stock trading
Answered June 19, 2017

Yes, definitely. It seems that after several changes Yahoo Finance closed their API forever. I think, it's a part of cost cutting strategy from Verizon. The API was closed on May 15, one month ago.

I've tested several alternatives and found that https://eodhistoricaldata.com the best one for those who used Yahoo Finance. The services provides raw data, adjusted closes and splits/dividends. They also have CSV output, with very similar format for Yahoo Finance users.

Also there is a https://intrinio.com/ data provider, looks good, but they much more expensive, have no data for Mutual Funds and API is very different in compare to Yahoo Finance. Then you need to significantly change your code.

24.5K views · View 5 upvotes

5 · 2

Can the Yahoo! Finance API be used for a commercial app? Is there any other free stock feed API that can be used in a commercial app?

Answer · Follow · 23 · Request

14 Answers

Adrian Ho, surgically removes semicolons; drinks @ spacebars
Answered September 4, 2015

Caveat: I'm in the process of rolling out my own securities analysis service, so I've some skin in this game, and actually talked to some of the players involved. It's possible that there are some corner cases that may invalidate some aspects of my analysis below, but it's not likely.

As I understand it, the Finance API is only for personal use. You aren't allowed to retrieve data for display on *any* app (web or otherwise) for third-party use for *any* purpose (free or commercial). That would be *redistribution*, and is expressly forbidden in the text at the bottom of every Finance page:

> Quotes are real-time for NASDAQ, NYSE, and NYSE MKT. See also delay times for other exchanges. All information provided "as is" for informational purposes only, not intended for trading purposes or advice. Neither Yahoo! nor any of independent providers is liable for any informational errors, incompleteness, or delays, or for any actions taken in reliance on information contained herein. **By accessing the Yahoo! site, you agree not to redistribute the information found therein.**

Clause 1.7.4. of the Yahoo APIs Terms of Use (Yahoo Developer Network Terms of Use) also applies:

> 1. Licensed Uses and Restrictions
> [...]
> 7. YOU SHALL NOT:
> [...]
> 4. Sell, lease, share, transfer, or sublicense the Yahoo APIs or access or access codes thereto or **derive income from the use or provision of the Yahoo APIs, whether for direct commercial or monetary gain or otherwise**, unless the API Documents specifically permit otherwise or Yahoo gives prior, express, written permission

I also don't know of any other free stock APIs that permits redistribution and/or commercial use, nor do I expect there will ever be one. These restrictions actually originate from the data providers (Thomson Reuters, Interactive Data, etc.); Yahoo! is paying the them for the right to display the data on their site and make it available for personal use.

That said, Yahoo! is probably unable to stop you from using the data against their policies. However, since the data can only be publicly used when acquired from a handful of specific sources, it's trivial for the corresponding data providers to begin investigating when they don't see you in their customer lists...and threaten a lawsuit when they uncover the truth.

11.6K views · View 7 upvotes · View shares

7 · 1 · 1

Table C3. Viewership for Quora Q&As related to Yahoo! Finance APIs

This table reports the number of views of the highest viewed answers to sample questions on Quora related to Yahoo! Finance APIs and its alternatives. The "Yahoo" entry is 1 if the question mentions or is about Yahoo! Finance API, and 0 otherwise. The number of views is as of March 25, 2021. We observe that viewers are generally interested in using the APIs to get stock quotes, why the API was shut down, and alternatives to the API. To put the numbers in perspective, the "Stock Market Traders" group on Quora has approximately 318,600 followers, which is roughly 2.5 times the highest number of views among the sample questions.

Year	Question	Yahoo	# of views
4-Nov-17	Did Yahoo Stock API shut down?	1	42500
29-Sep-17	What is the best alternative to Yahoo Finance?	1	18100
4-Sep-15	Can the Yahoo! Finance API be used for a commercial app? Is there any other free stock feed API that can be used in a commercial app?	1	11900
19-Jun-17	Where can I find Yahoo Finance API documentation?	1	11700
13-Apr-18	Why were the Google and Yahoo Finance APIs shutdown? Where and how did they get their data?	1	11500
19-Jun-17	Why did Yahoo discontinue it's finance API?	1	5600
4-Oct-19	What sites offer stock market information for free (ideally via API) similar to Yahoo?	1	3100
22-Nov-17	Why should we use Quandl instead of Yahoo API for historical real-time stock prices?	1	2600
9-Jun-17	As you know Yahoo! Finance API is broken for 3 weeks already and seems that Yahoo will no fix it. Let's share alternatives here?	1	2100
2-Oct-18	How can you use Yahoo! Finance API for CSV?	1	1900
7-Oct-18	How do I find the risk-free rate of a company on Yahoo Finance or any other sources?	1	1400
4-Jan-18	What is a free finance API that replaces the old Yahoo API for things like PEG ratio, EPS, etc. for Google spreadsheets?	1	1100
5-Jan-18	What is a free finance API that replaces the old Yahoo API for things like PEG ratio, EPS, etc. for Google spreadsheets?	1	1100
18-Sep-18	How up to date is the Yahoo Finance API?	1	848
16-Oct-16	I want my Android app to fetch data from Yahoo finance Api and refresh it on a 5 - 10 minute basis. How should I go about it?	1	765
1-Oct-18	What is the API URL for Yahoo finance Canada?	1	702
23-Nov-17	With Yahoo bailing, are there services today (Nov. 2017) that still offer (through curl or API) delayed market index and stock data for free?	1	573
12-Jun-18	What are some alternative free data sources for intra-day stock quotes, given that Google and Yahoo have recently ended these data services?	1	564
2-Jul-17	Have you ever built an app that uses the Yahoo! Finance API? What did it do?	1	418
12-May-20	What are Yahoo Finance API alternatives?	1	257
18-Jan-19	Using Python and Yahoo Finance API, there is an extra column called "Adj Close". How is this different from the "Close" column?	1	99
2-Nov-17	Is there a real time stock market data feed API for NSE, BSE, & Mcx to implement in our custom software?	0	132300
9-May-16	What are some good APIs to get real time stock quotes?	0	130400
31-Aug-18	What open source APIs can I use to get financial data automatically?	0	56800
11-Jun-18	Stock Market: Which Python libraries can I use to access stock market data in real time?	0	33800
7-Jul-18	What's the best free API for programmatically retrieving current stock price data?	0	28100
5-May-16	What are some good APIs for stock exchanges data?	0	14700
7-Jan-17	Which APIs provide real time data of BSE/NSE stock prices?	0	8600

4-Nov-18	How's alpha vantage comparing to Google finance API?	0	5600
30-Jul-15	Does Google Finance allow web scraping of data from its website?	0	5300
5-Jul-19	How can I get stock quotes using Google Finance API?	0	4600
7-Sep-17	Why is Google's finance API not working?	0	4300
30-Oct-18	Where do free financial data services, like Alpha Vantage, get their data?	0	2400
5-Jul-19	Are there any free or cheap Stock market APIs for commercial use?	0	2300
6-Sep-17	Has the Google Finance “get quote” API stopped working?	0	2300
15-Aug-18	How can we fix Google Finance API issues?	0	2100
17-Jul-19	Where can I find a free finance API to fetch company financials such as profit for Europe listed stocks?	0	2000
22-Apr-17	Why would I use a website API to get stock quotes instead of simply getting them from the NASDAQ or Yahoo Finance?	0	1900
14-Feb-18	How do I get every second currency exchange rates via an API (such as Google Finance)?	0	1900
14-Aug-18	How do I get stock market APIs?	0	1600
6-Apr-13	Algorithmic Trading: Is there a license free data source for historical stock prices?	0	1400
28-Jul-14	Where can I find a free (or affordable) finance API for Balance sheets, Income statements and Cash flow?	0	1300
8-Nov-17	Have you used Alpha Vantage's APIs? If so, what has been your experience with them?	0	917
10-Dec-17	How much longer will the Google Finance RESTFUL API function?	0	838
24-Jul-17	Are there any free APIs for checking changes in the major sectors of the stock market, i.e. the one that is displayed on the Google Finance page?	0	318
10-Jul-18	Where is the best place to find financial APIs?	0	229
24-Sep-20	What are some good stock APIs?	0	156
3-Jul-20	Are there any free stock APIs that provide a way to get the top companies by market cap?	0	38

Appendix D. Sample reactions to Yahoo! Finance API shutdown on Yahoo! Finance Help Community

The following screenshots are captured from the Wayback Machine archive of the discussion about the Yahoo! Finance API shutdown on Yahoo! Finance Help Community[17]. The full discussion has 25 pages.

GYang14 Established 'Hoo

✔ Is Yahoo! Finance API broken?

I was trying to use R quantmod to get historical stock price from Yahoo but get the error saying cannot open URL 'https://ichart.finance.yahoo.com/...'. When I manually type the address in the brower, I got following page https://ichart.finance.yahoo.com/ saying that "Yahoo will be right back... Thank you for your patience.....". Does this issue happen before? Will it be really solved in a short time?

Thank you very much.

Solved! Go to Solution.

17 Kudos

ADMIN Nixon Administratin' 'Hoo

Re: Is Yahoo! Finance API broken?

Hi All - I have reported this to the Finance engineering team. They will investigate and I will provide an update when I hear back. Sorry for the inconvenience!

If you see any helpful posts, give a Kudo! If you see a response that answers your question, please mark Accept as Solution!

9 Kudos

ADMIN Nixon Administratin' 'Hoo

✔ Re: Is Yahoo! Finance API broken?

Hi All - This feature was discontinued by the Finance team and they will not be reintroducing that functionality.

The Yahoo Finance Feedback Forum is the place where you can make product suggestions and provide feedback. We're always trying to improve our products and use your feedback to inform changes. Here's the url: https://yahoo.uservoice.com/forums/382977

If you see any helpful posts, give a Kudo! If you see a response that answers your question, please mark Accept as Solution!

4 Kudos

[17] https://web.archive.org/web/20171108022152/https://forums.yahoo.net/t5/Yahoo-Finance-help/Is-Yahoo-Finance-API-broken/td-p/250503

NICKITA Established 'Hoo

Re: Is Yahoo! Finance API broken?

This is literally an abuse of entire community! Have you thought about people who's software is relying on that API? There has been no **clear** notification given in advance. I just realised my app simply doesn't work because Yahoo closed its historical API. And now entire system has to be redesigned. Unbelievable. Okay, apart from complaining, can anybody advice another free historical service available? I'm just shocked it happened. Hours were wasted because landing page says "Our engineers are working quickly to resolve the issue." rather than "This API is not active anymore".

19 Kudos

dmunday New 'Hoo

Re: Is Yahoo! Finance API broken?

Shame about this. I've been trying to work around this for hours today but seems it's not easily possible. Shame as I've been using this data since 2003!! for my own use. I really don't see why they're evening providing the data at all anymore as nobody is really going to scroll through pages of this stuff, or save away file after file manually.Being able to make a single simple request to get the data was great.

Oh well. I've grown less found of Yahoo and my final reason to continue accessing the site has been removed.
I've also been using another service for a while but unfortunately only daily data available.

1 Kudo

NICKITA Established 'Hoo

Re: Is Yahoo! Finance API broken?

Thanks to everyone who is participating in this thread!

I'm glad there are people who understand that such an attitude to customers is **really** unacceptable.

After hours of browsing and research, I found that so far there is no such a service to provide same functionality as Yahoo FInance API.

BUT. If your application requires historical data in a rough format of:

Date,Open,High,Low,Close,Volume

Google provides historical service which I tested today and everything was working fine:

```
http://www.google.com/finance/historical?q=NASDAQ:ADBE&startdate=Jan+01%2C+2009&enddate=Aug+2%2C+2012&output=csv
```

I hope wrapping request in code block bypasses yahoo link block and allows you to test it out.

Result of request is:

```
Date,Open,High,Low,Close,Volume
2-Aug-12,30.46,30.83,30.25,30.58,2639980
1-Aug-12,31.05,31.17,30.64,30.69,2738740
31-Jul-12,30.99,31.27,30.84,30.88,2662853
...
```

I know, you may require much broader functionality, however hope for some of you it may be a temporary solution.

5 Kudos

Appendix E. Variable Definitions and Data Sources

Variables	Definitions	Data Sources
Retail_Vol	Shares of trades initiated by retail investors, scaled by total shares outstanding and multiplied by 100. Retail trades are identified based on TAQ exchange code (D) and a small price improvement (0-0.4 cents, exclusive, above (below) a round cent for sale (buy) transactions), following Boehmer et al. (2021).	TAQ & CRSP
Ab_Retail_Vol	*Retail_Vol* minus its median for the same day of the week over the past 10 weeks.	TAQ & CRSP
Ab_Retail_Buy	Shares of trades bought by retail investors (scaled by total shares outstanding and multiplied by 100) minus its median for the same day of the week over the past 10 weeks.	TAQ & CRSP
Ab_Retail_Sell	Shares of trades sold by retail investors (scaled by total shares outstanding and multiplied by 100) minus its median for the same day of the week over the past 10 weeks.	TAQ & CRSP
AIM	Amihud(2002) illiquidity measure, the natural logarithm of the ratio of absolute stock return to dollar volume [$1{,}000{,}000 \times \| ret \| \div (prc \times vol)$]	CRSP
Spread	Daily bid-ask spread based on CRSP data, $100 \times (ask - bid)/[(ask + bid)/2]$.	CRSP
CAR[1W]	Cumulative abnormal return for next week, starting from the next (trading) day.	CRSP
CAR[iW, jW]	Cumulative abnormal return from week i to week j, both inclusive ($i \neq 0$)	CRSP
Institutional_Vol	Shares of trades initiated by institutional investors, scaled by total shares outstanding and multiplied by 100. Institutional trades are non-retail trades with trade sizes above $50,000, following Bushee et al. (2020).	TAQ & CRSP
Ab_Institutional_Vol	*Institutional_Vol* minus its median for the same day of the week over the past 10 weeks.	TAQ & CRSP
RFS	Retail-favored stock. Equal to 1 if the stock's retail holding (measured as shares not reported to be held by institutional investors, scaled by total shares outstanding) is above sample median, 0 otherwise.	Thomson Reuters
Low-priced	Equal to 1 if the stock price is below the sample median, 0 otherwise. The stock price is measured as of the last trading day before the main sample starts.	CRSP
High-volatility	Equal to 1 if the idiosyncratic stock volatility is above sample median, 0 otherwise. The volatility is measured as the standard deviation of the residual returns from estimating a four-factor model on daily returns during the six months (October 16, 2016 to April 15, 2017) before the starting date of the sample, following Kumar (2009).	CRSP
High-skewness	Equal to 1 if the stock skewness is above sample median, 0 otherwise. Skewness is defined as the third moment of the residual obtained from estimating a two-factor model on daily returns during the six months (October 16, 2016 to April 15, 2017) before the starting date of the sample, following Kumar (2009).	CRSP
Lottery-like	Equal to 1 if *Low-priced* = *High-volatility* = *High-skewness* = 1, 0 otherwise.	CRSP
Small-cap	Equal to 1 if the stock market capitalization is below the sample median, 0 otherwise. We measure the market capitalization as of the last trading day before the main sample starts.	CRSP
Post	Equal to 1 if the date is on or after May 16, 2017, 0 otherwise.	-
Ret	Delist adjusted stock returns.	CRSP
Ret^2	Square of *Ret*.	CRSP
News	Natural logarithm of one plus the number of news articles on the Dow Jones Edition of RavenPack with relevance store above 20 (the company name can be identified somewhere in the story).	RavenPack
Size	Natural logarithm of market capitalization (*prccq**cshoq*) at the fiscal year-end.	Compustat
ROA	Return on assets (*ib/at*).	Compustat
Loss	Equals to 1 if *ROA*<0, 0 otherwise	Compustat
R&D	R&D intensity (*xrd/at*).	Compustat

Advertising	Advertising intensity (*xad/at*).	Compustat
Leverage	Financial leverage ((*dltt+dlc)/at*).	Compustat
Analysts	Natural logarithm of one plus the number of financial analysts following the company.	IBES
BTM	Book to market ratio, measured as the ratio of the book value of the equity to its market value.	Compustat & CRSP

Appendix F. Analyses on Sub-indices of Lottery-like Stocks

This appendix presents analyses using sub-indices of lottery-like stocks, as alternative proxies for retail-favored stocks. The three indices are low-priced, high-volatility, and high-skewness following Kumar (2009). Analyses using the composite index of lottery-like stocks are presented in the main manuscript.

Figure F1. Retail trading volume around the shutdown of Yahoo! Finance API

This figure plots daily retail trading volume around the shutdown of Yahoo! Finance API for lottery-like stocks (solid line) and other stocks (dashed line), respectively. The y-axis is retail trading volume scaled by total shares outstanding, multiplied by 100. Panels A-C designate stocks with low price, high volatility, and high skewness, respectively as alternative proxies for retail-favored stocks. The vertical dashed lines indicate the shut-down of Yahoo! Finance API.

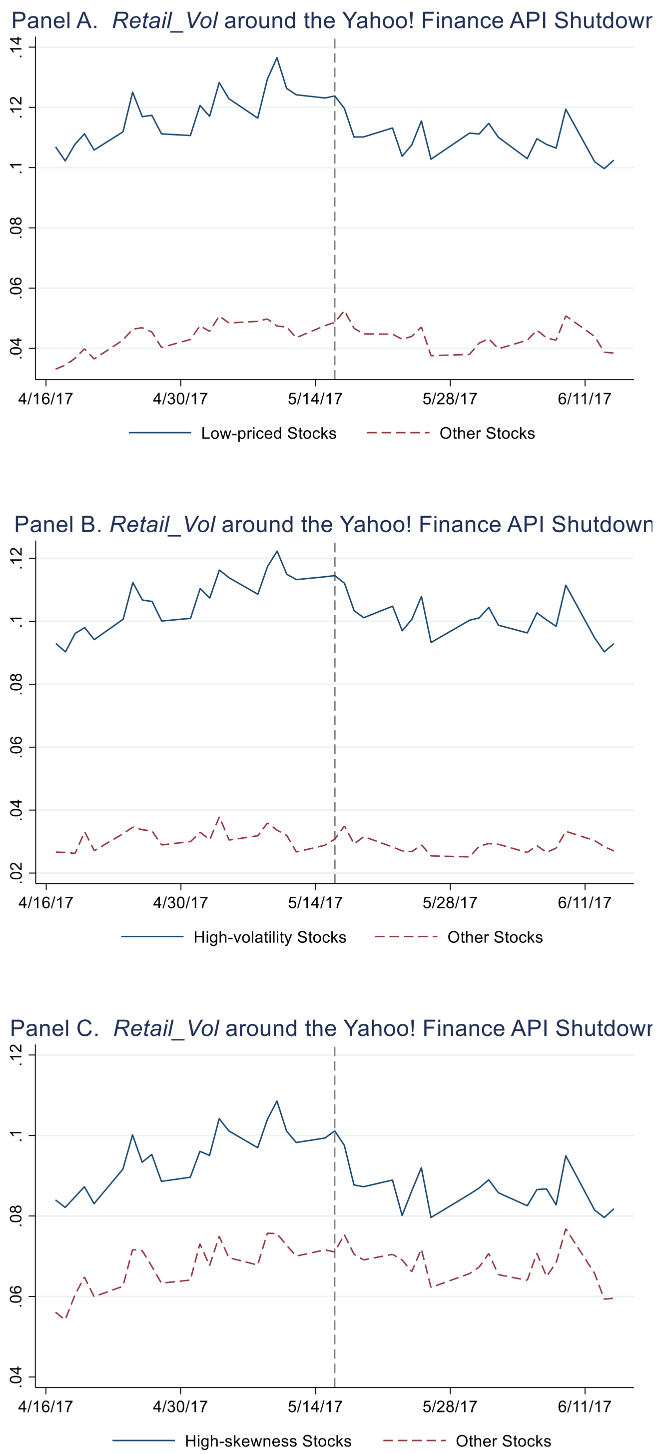


Table F1. Univariate Comparisons

	Pre	Post	MeanDiff	Pre	Post	MeanDiff	DID
		Low_Priced = 1			**Low_Priced = 0**		
Retail_Vol	0.121	0.112	-0.009***	0.044	0.044	0.000	-0.009***
Ab_Retail_Vol	0.028	0.023	-0.005***	0.012	0.011	-0.001	-0.005*
		High_Volatility = 1			**High_Volatility = 0**		
Retail_Vol	0.109	0.103	-0.006***	0.031	0.029	-0.003***	-0.004*
Ab_Retail_Vol	0.027	0.023	-0.004*	0.007	0.005	-0.002***	-0.001
		High_Skewness = 1			**High_Skewness = 0**		
Retail_Vol	0.097	0.089	-0.008***	0.069	0.068	-0.000	-0.008***
Ab_Retail_Vol	0.023	0.018	-0.005***	0.018	0.018	0.000	-0.006**

Table F2. Retail trades around the shutdown of Yahoo! Finance API

This table reports the regression results of retail trades around the shutdown of Yahoo! Finance API (May 16, 2017). The sample is a panel of firm-days in a two-week, two-, or four-month window (indicated in the table header) centered around the shutdown. The dependent variable is the abnormal retail trading volume (*Ab_Retail_Vol*). The key variable of interest is the interaction between *Post* and the treatment group, i.e., stocks with lottery-like features such as stocks with low price (*Low_Priced*), high volatility (*High_Volatility*), and high skewness (*High_Skewness*) as indicated in the table header. See Appendix E for detailed variable definitions. Robust standard errors clustered by firm are reported in the parentheses. ***, **, and * stand for statistical significance at the 1%, 5%, and 10% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treat =		Low_Priced			High_Volatility			High_Skewness	
	2-Week	2-Month	4-Month	2-Week	2-Month	4-Month	2-Week	2-Month	4-Month
Post×Treat	-0.008***	-0.005**	-0.002	-0.005**	-0.001	-0.001	-0.009***	-0.007***	-0.004*
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Ret	0.158***	0.133***	0.151***	0.158***	0.133***	0.151***	0.159***	0.132***	0.151***
	(0.046)	(0.027)	(0.020)	(0.046)	(0.027)	(0.020)	(0.046)	(0.027)	(0.020)
Ret^2	29.636***	33.313***	35.817***	29.650***	33.313***	35.818***	29.669***	33.320***	35.816***
	(1.087)	(0.814)	(0.821)	(1.088)	(0.814)	(0.821)	(1.087)	(0.814)	(0.821)
News	0.021***	0.022***	0.023***	0.021***	0.022***	0.023***	0.021***	0.022***	0.023***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	39,433	169,426	326,635	39,433	169,426	326,635	39,433	169,426	326,635
R-squared	0.558	0.362	0.291	0.558	0.362	0.291	0.558	0.362	0.291

Table F3. Placebo tests using institutional trades

This table reports the regression results of institutional trades around the shutdown of Yahoo! Finance API (May 16, 2017). The sample is a panel of firm-days in a two-month window centered around the shutdown. The dependent variable is the abnormal institutional trading volume (*Ab_Institutional_Vol*). The regression specification is the same as in Table 3. The key variable of interest is the interaction between *Post* (indicating the period after the shutdown of Yahoo! Finance API) and *Treat* the treatment group, i.e., stocks with lottery-like features such as stocks with low price (*Low_Priced*), high volatility (*High_Volatility*), and high skewness (*High_Skewness*) as indicated in the table header. See Appendix E for detailed variable definitions. Robust standard errors clustered by firm are reported in the parentheses. ***, **, and * stand for statistical significance at the 1%, 5%, and 10% level, respectively.

DV = *Ab_ Institutional _Vol*	(1)	(2)	(3)
Treat =	Low_Priced	High_Volatility	High_Skewness
Post×Treat	-0.001	0.002	-0.001
	(0.002)	(0.002)	(0.002)
Ret	0.041**	0.041**	0.041**
	(0.021)	(0.021)	(0.021)
Ret^2	18.849***	18.850***	18.849***
	(0.626)	(0.626)	(0.626)
News	0.037***	0.037***	0.037***
	(0.001)	(0.001)	(0.001)
Day FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	169,211	169,211	169,211
R-squared	0.194	0.194	0.194

Table F4. Falsification tests

This table reports the results of falsification tests in a two-month window centered around May 16, 2016, one year before the shutdown of Yahoo! Finance API. The dependent variable is the abnormal retail trading volume (*Ab_Retail_Vol*). The regression specification is the same as in Table 3. The key variable of interest is the interaction between *Post* (indicating the period after the shutdown of Yahoo! Finance API) and *Treat* the treatment group, i.e., stocks with lottery-like features such as stocks with low price (*Low_Priced*), high volatility (*High_Volatility*), and high skewness (*High_Skewness*) as indicated in the table header. See Appendix E for detailed variable definitions. Robust standard errors clustered by firm are reported in the parentheses. ***, **, and * stand for statistical significance at the 1%, 5%, and 10% level, respectively.

DV = *Ab_ Retail _Vol*	(1)	(2)	(3)
Treat =	Low_Priced	High_Volatility	High_Skewness
Post×Treat	0.000	0.001	-0.002
	(0.002)	(0.001)	(0.002)
Ret	0.026	0.026	0.027*
	(0.016)	(0.016)	(0.016)
Ret^2	21.348***	21.350***	21.348***
	(0.506)	(0.506)	(0.506)
News	0.020***	0.020***	0.020***
	(0.001)	(0.001)	(0.001)
Day FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	156,593	156,593	156,593
R-squared	0.362	0.362	0.362

Table F5. The predictivity of retail trades for future returns

This table reports the daily Fama-Macbeth regressions of future returns on abnormal retail buy and sell trading volume. The sample includes firm-day observations during a 2-month window centered around the Yahoo! Finance API shutdown. We run separate analyses in lottery-like stocks (low price, high volatility, and high skewness) and other stocks. See Appendix E for detailed variable definitions. Newey and West (1987) standard errors with lags of two are reported in the parentheses. ***, **, *, and # stand for statistical significance at the 1%, 5%, 10%, and 12% level, respectively.

DV = Sample	CAR[1W,4W] & Low_Price=0		CAR[1W,4W] & Low_Price=1	
	Pre	Post	Pre	Post
Ab_Retail_Buy	6.504*	4.744*	-0.321	5.432***
	(3.225)	(2.707)	(1.928)	(1.261)
Ab_Retail_Sell	-1.367	-10.107***	-5.015***	-5.628***
	(3.684)	(2.926)	(1.120)	(1.341)
Ab_Retail_Buy (Col 4-Col 3) – (Col 2-Col 1):		7.513#		
Ab_Retail_Sell (Col 4-Col 3) – (Col 2-Col 1):		8.127#		
Controls	Yes	Yes	Yes	Yes
Observations	38,279	39,958	44,507	46,388
R-squared	0.029	0.050	0.016	0.012
DV = Sample	CAR[1W,4W] & High_Volatility=0		CAR[1W,4W] & High_Volatility =1	
	Pre	Post	Pre	Post
Ab_Retail_Buy	3.318	-9.155***	0.466	5.585***
	(2.816)	(3.206)	(1.785)	(1.285)
Ab_Retail_Sell	1.216	-7.122**	-4.541***	-6.060***
	(4.088)	(2.958)	(1.086)	(1.211)
Ab_Retail_Buy (Col 4-Col 3) – (Col 2-Col 1):		17.592***		
Ab_Retail_Sell (Col 4-Col 3) – (Col 2-Col 1):		6.819		
Controls	Yes	Yes	Yes	Yes
Observations	25,191	26,277	57,595	60,069
R-squared	0.061	0.070	0.017	0.013
DV = Sample	CAR[1W,4W] & High_Skewness=0		CAR[1W,4W] & High_Skewness=1	
	Pre	Post	Pre	Post
Ab_Retail_Buy	4.386	2.750	-0.970	6.193***
	(3.928)	(3.106)	(1.828)	(1.116)
Ab_Retail_Sell	-7.005***	-7.195***	-3.027*	-5.323***
	(2.411)	(2.212)	(1.546)	(1.340)
Ab_Retail_Buy (Col 4-Col 3) – (Col 2-Col 1):		8.799#		
Ab_Retail_Sell (Col 4-Col 3) – (Col 2-Col 1):		-2.106		
Controls	Yes	Yes	Yes	Yes
Observations	33,323	34,773	49,463	51,573
R-squared	0.021	0.032	0.025	0.016

Table F6. Market liquidity

This table reports the regression results of market liquidity in a two-month window centered around the shutdown of Yahoo! Finance API. The dependent variable is the daily Amihud's illiquidity measure (*AIM*) in Panel A and the daily relative bid-ask spread (*Spread*) in Panel B. The key variable of interest is the interaction between *Post* (indicating the period after the shutdown of Yahoo! Finance API) and *Treat* the treatment group, i.e., stocks with lottery-like features such as stocks with low price (*Low_Priced*), high volatility (*High_Volatility*), and high skewness (*High_Skewness*) as indicated in the table header. . See Appendix E for detailed variable definitions. Robust standard errors clustered by firm are reported in the parentheses. ***, **, and * stand for statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A. AIM

DV = *AIM*	(1)	(2)	(3)
Treat =	Low_Priced	High_Volatility	High_Skewness
Post×Treat	0.010***	0.008***	0.005**
	(0.002)	(0.002)	(0.002)
Ret	-0.218***	-0.217***	-0.218***
	(0.025)	(0.025)	(0.025)
Ret^2	16.131***	16.133***	16.125***
	(0.727)	(0.727)	(0.727)
News	-0.012***	-0.012***	-0.012***
	(0.001)	(0.001)	(0.001)
Day FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	168,917	168,917	168,917
R-squared	0.612	0.612	0.612

Panel B. Spread

DV = *Spread*	(1)	(2)	(3)
Treat =	Low_Priced	High_Volatility	High_Skewness
Post×Treat	0.038***	0.024***	0.011
	(0.009)	(0.007)	(0.009)
Ret	0.148*	0.150*	0.148*
	(0.089)	(0.089)	(0.089)
Ret^2	13.109***	13.111***	13.091***
	(1.774)	(1.774)	(1.774)
News	-0.010***	-0.010***	-0.009***
	(0.002)	(0.002)	(0.002)
Day FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	169,424	169,424	169,424
R-squared	0.753	0.753	0.753

Appendices G. Additional Robustness Checks

We further differentiate retail trading volumes in Table 3 to buy and sell volumes for retail trades. Abnormal retail buys (sells) are the retail buy (sell) trading volume scaled by shares outstanding, and minus its median value over the last 10 weeks. Institutional buy and sell volumes are similarly defined. Using the same predictors as in Table 2, we find that retail buys and sells drop by a similar magnitude after the shutdown of Yahoo! Finance API (Table G1).

As a robustness test, we assess whether the results are sensitive to the control and treatment group classification. We replicate the main analyses while excluding sample firms whose institutional holdings fall in the middle 20 percent. Effectively, we changed the treatment and control groups from those with below and above median institutional holdings to 0-40% and 60%-100%, respectively. The results in Table G2 are consistent with that of Tables 3 and 6.

Table G1: Buy versus sell trades

This table reports the impact of Yahoo! Finance API shutdown on retail buy and sell volumes, respectively. The sample period is the two-month window centered around May 16, 2017. The dependent variables are abnormal retail trading buy and sell (*Ab_Retail_Buy, Ab_Retail_Sell*), indicated by "Buy" and "Sell" in the table header, respectively. The key variable of interest is the interaction between *Post* (indicating the period after the shutdown of Yahoo! Finance API) and *Treat* (stocks preferred by retail investors based on their holdings (*RFS*), lottery-like stocks (*Lottery_like*), or small-cap stocks (*Small_Cap*), as indicated in the table header). See Table A1 for detailed variable definitions. Robust standard errors clustered by firm are reported in the parentheses. ***, **, and * stand for statistical significance at the 1%, 5%, and 10% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Treat=	*RFS*		*Lottery-Like*		Small_Cap	
DV=	Buy	Sell	Buy	Sell	Buy	Sell
Post×Treat	-0.003***	-0.003***	-0.005***	-0.004***	-0.003**	-0.003**
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	169,430	169,430	169,426	169,426	169,426	169,426
R-squared	0.345	0.337	0.345	0.337	0.345	0.337

Table G2: Alternative sample construction: excluding the middle 20%

Panel A and B of this table are robustness checks for Table 3 Panel A and B, and Table 6 Panel B, respectively. The regression specifications are the same as before. The only difference is that we exclude firms whose retail holdings in the middle 20 percentile (above 40th percentile and below 60th percentile) from the analyses. See Appendix E for detailed variable definitions. Robust standard errors clustered by firm are reported in the parentheses. ***, **, and * stand for statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: Retail trades

	(1)	(2)	(3)	(4)	(5)	(6)
DV= *Ab_Retail_Vol*	**2-Week Window**		**2-Month Window**		**4-Month Window**	
Post×RFS	-0.011***	-0.011***	-0.010***	-0.010***	-0.003	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Ret	0.168***	0.184***	0.130***	0.159***	0.170***	0.184***
	(0.062)	(0.049)	(0.033)	(0.030)	(0.024)	(0.023)
Ret^2	37.870***	30.653***	36.743***	33.833***	37.574***	36.266***
	(1.536)	(1.196)	(1.097)	(0.906)	(1.016)	(0.903)
News	0.023***	0.024***	0.020***	0.024***	0.021***	0.026***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
RFS	0.004		0.001		-0.003	
	(0.004)		(0.002)		(0.002)	
Size	-0.002		-0.001		-0.001	
	(0.001)		(0.001)		(0.001)	
ROA	0.001		0.001		0.002	
	(0.003)		(0.002)		(0.002)	
Loss	0.007		-0.006		-0.017*	
	(0.014)		(0.010)		(0.009)	
R&D	0.000		-0.004		-0.006***	
	(0.004)		(0.003)		(0.002)	
Advertising	-0.033		-0.026		-0.017	
	(0.022)		(0.017)		(0.014)	
Leverage	0.335***		0.224***		0.115***	
	(0.107)		(0.066)		(0.043)	
Analysts	0.009		0.002		-0.000	
	(0.008)		(0.005)		(0.003)	
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	No	Yes	No	Yes	No
Firm FE	No	Yes	No	Yes	No	Yes
Observations	34,728	34,728	135,673	135,673	261,592	261,592
R-squared	0.233	0.541	0.212	0.360	0.215	0.289

Panel B. Return Analyses

	(1)	(2)	(3)	(4)
DV=	CAR[1W,4W] &		CAR[1W,4W] &	
Sample	RFS=0		RFS=1	
	Pre	Post	Pre	Post
Ab_Retail_Buy	8.630**	-3.234	-0.815	7.180***
	(3.476)	(2.760)	(1.566)	(1.456)
Ab_Retail_Sell	-0.362	-4.523	-5.273***	-5.797***
	(4.634)	(3.359)	(1.119)	(1.247)
Ret[0]^2	-288.990***	169.891	-277.427***	-78.539
	(96.757)	(116.482)	(61.125)	(50.407)
News	0.217	0.234**	0.221*	0.008
	(0.131)	(0.102)	(0.116)	(0.128)
Size	0.399***	-0.429***	0.219***	0.193***
	(0.109)	(0.086)	(0.023)	(0.049)
BTM	-1.681***	0.037	0.674***	0.043
	(0.230)	(0.315)	(0.094)	(0.126)
Ret[0]	12.971**	-3.853	4.444	-15.067***
	(5.088)	(8.872)	(3.755)	(4.392)
CAR[-1W]	0.051	-0.016	0.076***	-0.044*
	(0.032)	(0.024)	(0.019)	(0.023)
CAR[-2W,-4W]	0.082***	-0.028	0.075***	0.014
	(0.019)	(0.025)	(0.017)	(0.011)
Ab_Retail_Buy (Col 4-Col 3) – (Col 2-Col 1):			19.859***	
Ab_Retail_Sell (Col 4-Col 3) – (Col 2-Col 1):			3.637	
Observations	33,291	34,763	32,681	34,116
R-squared	0.034	0.033	0.021	0.014
Number of groups	21	22	21	22