Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Tech-Enabled Financial Data Access, Retail Investors, and Gambling-Like Behavior in the Stock Market

Taha Havakhor; , Mohammad Saifur Rahman; , Tianjian Zhang; , Chenqi Zhu

informs
https://pubsonline.informs.org/journal/mnsc
MANAGEMENT SCIENCE
Vol. 71, No. 2, February 2025, pp. 1646–1670
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Tech-Enabled Financial Data Access, Retail Investors, and Gambling-Like Behavior in the Stock Market

Taha Havakhor,[a] Mohammad Saifur Rahman,[b,*] Tianjian Zhang,[c] Chenqi Zhu[d]

[a] McGill University, Montreal, Quebec H3A 0G4, Canada; [b] Purdue University, West Lafayette, Indiana 47907; [c] California State University, Dominguez Hills, Carson, California 90747; [d] University of California, Irvine, Irvine, California 92697

*Corresponding author

Contact: taha.havakhor@mcgill.ca, https://orcid.org/0000-0002-4338-5970 (TH); mrahman@purdue.edu, https://orcid.org/0000-0003-2115-5776 (MSR); tzhang@csudh.edu, https://orcid.org/0000-0003-1035-2838 (TZ); chenqiz1@uci.edu, https://orcid.org/0000-0003-1028-2374 (CZ)

Received: April 26, 2021
Revised: August 4, 2022; March 12, 2023
Accepted: May 21, 2023
Published Online in Articles in Advance: May 16, 2024

https://doi.org/10.1287/mnsc.2021.01379

Abstract. Advancements in technology have reduced information acquisition costs, creating an improved information environment for retail investors. Specifically, new technologies, such as application programming interface (API), deliver high-volume, institutional-like raw data directly to Main Street investors. Although greater availability of information can be beneficial, it may also exacerbate retail investors' existing trading deficiencies. Exploiting the sudden shutdown of Yahoo! Finance API, the largest free API for retail investors, this study examines how access to tech-enabled raw financial data affects retail investment. We find that retail trading volumes in stocks favored by active retail investors dropped by 8.6%–10.5% within one month of the API shutdown. The remaining retail trades collectively became more predictive of future returns, suggesting less gambling-like behavior after the API shutdown. Moreover, our randomized controlled experiment affirms the underlying mechanism: tech-enabled access to high-volume historical price data increases individuals' overconfidence, which further leads them to engage in excessive trading. The study reveals an unintended consequence of technology-led, wider data access for retail investors.

History: Accepted by D. J. Wu, information systems.

Supplemental Material: The online appendix and data files are available at https://doi.org/10.1287/mnsc.2021.01379.

Keywords: retail investors • financial technology • financial data • gambling • noise trading • stock market • application programming interface • quasi-natural experiment • randomized controlled experiment

1. Introduction

Retail (or Main Street) investors are individuals who invest in the stock market using their personal accounts. Their well-being and potential impact on financial markets have long been of interest to economists and regulators. The advent of financial technology (FinTech) has reduced the cost to conduct trades and access investment-related information, thus attracting more individuals to the stock market, and led to events with severe implications for retail investors[1] and the market as a whole.[2] With growing interest in the impact of FinTech on Main Street investors, we study whether and how financial data application programming interfaces (APIs), an understudied category of FinTech, affect retail trading. APIs can close the gap between Main Street and Wall Street investors in accessing large-scale market data. However, considering the unknown impacts of enhanced data access on retail investors' behavioral biases, it is ex ante unclear whether APIs are a boon or bane for retail investors.

Financial data APIs allow easy access to a large volume of raw data on demand (mostly historical and real-time price data) with which users can seamlessly implement and automate nimble trading strategies that incorporate the most updated data.[3] In today's competitive and fast-moving market, APIs are essential to gain an informational edge or not lose out, and institutional investors pay a high price to access APIs (e.g., Xignite, Bloomberg API). As the technology expands and the costs decline, a few low-cost or free APIs have become popular among retail investors. However, the benefits of providing them with an institutional-like data feed should be questioned given that Main Street investors are typically less sophisticated and more vulnerable to behavioral biases (Barber and Odean 2013). Prior research on tech-enabled information dissemination mainly focuses on the provision of low-frequency financial information, such as accounting information and investment research (Gao and Huang 2020, Farrell et al. 2022). As price data are constantly updated with high volumes, financial data APIs facilitate a fundamentally different type of information dissemination in terms of the type of information, scale of information transmission, and underlying technology (e.g., real-time and scalable delivery, which allows nimble and automated trading strategies). Against this backdrop, this study examines how financial technologies that facilitate high-volume dissemination of raw data affect retail investment.

Retail investors are often dubbed "noise traders" as they lack cost-effective access to information, have gambling-like trading preferences, and suffer from various behavioral biases. In addition to their inherent informational disadvantage, many retail investors focus on past price trends and ignore key fundamental information (Blankespoor et al. 2019). They often treat stock market investments as substitutes for gambling (Kumar 2009, Dorn et al. 2015, Gao and Lin 2015) and are overly confident (Barber and Odean 2007, Grinblatt and Keloharju 2009). Although technology can improve the information environment for retail investors (Gao and Huang 2020, Farrell et al. 2022, Lee and Zhong 2022), there is no guarantee it will mitigate their cognitive biases. Specifically, API-enabled access to a large amount of historical price data likely strengthens their reliance on past returns and negligence on fundamental variables. Whereas repeating old mistakes is already problematic, access to a large volume of price data may worsen the quality of retail trades as investors become more overconfident as reflected in their illusions of knowledge, precision, and control.[4] Specifically, playing with data to look for patterns may create a false sense of knowledge (illusion of knowledge), leading Main Street investors to believe that their trading ability is higher than that of the average person. More data may also create an inflated sense of certainty when making predictions about stock prices (illusion of precision). Personal involvement in pulling and analyzing the data may lead investors to overestimate their control over the investment outcome (illusion of control). Thus, financial technologies enabling large-volume price data may not fix behavioral biases, but may aggravate them. In general, exacerbated overconfidence lead to excessive and lower quality trades (e.g., Barber and Odean 2002).

Our empirical strategy combines a quasi-natural experiment and a randomized controlled experiment. The quasi-natural experiment provides stock-level findings indicating that, when given access to API-enabled financial data, retail investors trade excessively and to their detriment. The randomized controlled experiment provides individual-level evidence that API access inflates overconfidence, which, in turn, leads to excessive trading and worse performance. The two experiments complement each other with the quasi-natural experiment documenting the collective effect on retail investors and the randomized controlled experiment delineating the underlying mechanism at the individual level.

The quasi-natural experiment exploits the abrupt shutdown (on May 16, 2017) of the Yahoo! Finance API. Yahoo! Finance was the most popular financial resource for retail investors at the time Lawrence et al. (2017). Its free API is more appealing to retail investors who actively engaged in feedback trading (trades based on stock price trends) because historical price data were the most prominent accessible data through the API. The behavioral heuristics of these retail investors often lead them to make investments congregating on gambling stocks (stocks that offer a skewed payoff) as they view trading as a get-rich-quick scheme (Dorn et al. 2015, Gao and Lin 2015). Thus, the shutdown of Yahoo! Finance API, arguably, disproportionately affected this group of retail investors. Leveraging this clientele effect, we employ a difference-in-differences (DID) design and compare retail trades in retail-favored stocks (treatment group) with retail trades in other stocks (control group) before and after the API was shut down.

The results of the quasi-natural experiment show that the retail trading volume dropped by 8.6%–10.5% in retail-favored stocks in a two-month window centered around the shutdown of Yahoo! Finance API. The remaining retail trades as a whole can better predict future returns, suggesting that retail trades given API access (before the shutdown) were less informed on average. These results are consistent with the argument that API-enabled quick and convenient access to raw financial data increases retail investors' gambling-like behaviors. When we extend the window to four months, however, the decrease in retail trading volume and increase in return predictability become much smaller and statistically insignificant, suggesting that their gambling-like behaviors returned as retail investors likely switched to alternative data sources (e.g., other APIs). We also find that market liquidity deteriorated significantly after the API shutdown, which is consistent with prior evidence that retail investors are, on average, liquidity providers (Grossman and Miller 1988, Kaniel et al. 2008, Kelley and Tetlock 2013) and public information improves market liquidity and attracts uninformed trades (Han et al. 2016).

To confirm the validity of our findings, we rule out alternative explanations related to investor selection, firm shocks (e.g., unobservable shocks or confounding events that reduce overall trading for retail-favored stocks), and market seasonality (e.g., retail-favored stocks move into quiet sessions in June compared with May). We recognize that retail investors who use APIs share common traits with other typical retail investors in terms of an informational disadvantage as well as behavioral biases (e.g., a preference for price over fundamental information and overconfidence). Where these retail investors might differ is in their technical skills. However, this difference should not bias our DID estimates because investors' technical skills should be similar immediately before and after the API shutdown.[5] To address concerns related to unobserved confounders, we conduct a placebo test using institutional trades. If our results are indeed driven by confounding shocks to retail-favored stocks that are independent from but coinciding with the time of the Yahoo! API shutdown, institutional investors would also respond to such shocks based on economic theory. However, we find that institutional trades do not change significantly around the API shutdown. Additionally, we do not find significant changes in retail trading volume one year before the actual shutdown, suggesting that our results cannot be explained by retail trading seasonality.

Our observational analysis of the quasi-natural experiment provides large-scale evidence, but the aggregated nature of stock-level analyses limits our ability to substantiate the underlying mechanism through which API access influences trading. To affirm the mechanism, we conduct a randomized controlled experiment to study whether access to tech-enabled financial data changes retail investors' overconfidence levels. We randomly expose subjects to API-like access to financial data. Consistent with our findings in the quasi-natural experiment, we observe that these individuals invest more in the stock market and incur greater losses. More importantly, we find that broad access to data leads to investors' overconfidence as reflected in an overestimation of their investment knowledge, an inflated sense of accuracy in predictions, and a false sense of control in trading. Paired with findings in the quasi-natural experiment, the randomized experiment indicates that access to tech-enabled financial data increases investors' overconfidence, which impacts the volume and quality of retail trading.

This study adds to the literature on tech-enabled information provision for retail investment. Prior research studies the provision of retail-grade information through brokerages (Barber and Odean 2002, Frydman and Wang 2020), trading applications (Liao et al. 2021, Barber et al. 2022), digitized corporate filings (Gao and Huang 2020), and crowdsourcing platforms (Ammann and Schaub 2021, Farrell et al. 2022). As Online Appendix A shows, our study focuses on the consequences of providing institutional-like information to retail investors, which includes a large amount of real-time and historical price data that are less processed and more dynamic in nature. We show that technologies disseminating and automating such data to retail investors aggravate their flawed trading instincts by intensifying retail traders' overconfidence, thus highlighting the potential drawbacks of tech-enabled information provision. The unintended consequences of market data democratization revealed in this study complement existing studies that largely document the benefit of providing tech-enabled information to retail investors.

This study also relates to the broader literature on retail investors' adoption of financial technologies. Prior studies show that online trading and trading applications that reduce transaction costs lead to excessive trading and worse overall performance by retail investors (e.g., Barber and Odean 2002, Ozik et al. 2021). We extend this literature by showing that data-democratizing financial technologies—which, in theory, are only welfare-enhancing for fully rational investors—can have similar adverse effects for retail investors. Some prior studies (e.g., Barber and Odean 2001) suggest that overconfidence drives their findings in retail investments. We provide experimental evidence that directly measures this behavioral bias in the presence of financial technology. Relatedly, our study further highlights the importance of elevating retail investors' awareness of their behavioral biases, echoing the U.S. Securities and Exchange Commission's (SEC) call to educate this group of investors (U.S. Securities and Exchange Commission 2012).

2. Institutional and Theoretical Background

2.1. Financial Data APIs

Financial data APIs provide users with large amounts of real-time, historical, and key reference data at lightning speed. Financial institutions typically pay hefty fees to obtain live data feeds in the form of APIs from stock exchanges or third-party sources (e.g., Xignite, Bloomberg's Server API).[6] For retail investors, free alternatives partially eliminate this price barrier. For example, Yahoo! Finance and Google Finance each offered a free API that allowed users to download batches of financial data with a simple query. Other examples include Alpha Vantage, Intrinio, and Tiingo. These low-cost or free financial data APIs are less extensive than institutional sources but offer retail investors cost-effective alternatives to access large amounts of financial data.

Yahoo! Finance API was the most popular financial data API among retail investors at the time. Although the exact number of API users is not publicly known, based on the volume of Google searches, Yahoo! Finance

API was the most searched compared with other alternatives (Online Appendix B). Its popularity was perhaps partially because of its popular Yahoo! Finance website, which was the most visited website for financial information, attracting more than 30 million unique daily users in 2016 (Lawrence et al. 2017). Other notable reasons that contributed to its popularity included Yahoo's extensive database of market transactions over the years, the relative reliability, and the free-access model.

Yahoo! Finance API (https://ichart.finance.yahoo.com) offered access to real-time data (e.g., real-time bid and ask price, last trade price), historical data (e.g., open and close price), and a few basic accounting variables (e.g., earnings per share, revenue, and book value), but no other commonly used ratios, such as return on assets (ROA), gross profit margin, or liquidity.[7] Online Appendix C exhibits an extended list of variables that were accessible through the Yahoo! Finance API. Although we cannot observe which variables were more frequently requested, most users seemed to be interested in historical price data (Online Appendices D and E). As one retail investor mentioned in the Yahoo! Finance Help Community,[8] "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."

Platforms such as YouTube, Stack Overflow, and Quora provided tutorials and tips on how to access historical price data through Yahoo! Finance API using software (Online Appendix D). For example, some YouTube tutorials laid out step-by-step guides and ready-to-use codes for obtaining stock quotes. Users with more advanced technical skills connected data analytics software, such as MATLAB, Mathematica, and R, with the Yahoo! Finance API. Additionally, some third-party applications facilitating feedback trading and technical analyses (e.g., Amibroker)[9] relied on Yahoo! Finance API for historical data. As such, the API generated a community of users with various levels of technical skills, and it functioned as a critical gateway for retail investors who heavily relied on historical price data.

These anecdotes suggest that the key value of Yahoo! Finance API was the efficient, scalable, and real-time access to large amounts of price data. Although some fundamental variables were accessible through the API, the price data (by nature) were updated far more frequently than the fundamental variables (daily or five-minute versus quarterly updates). Thus, the API was more convenient with greater utility when used to obtain high-volume historical price data or to frequently (or in real time) update the variables used in trading strategies. Importantly, Yahoo! Finance API provided a scalable tool for stock picking that differs from simply browsing through Yahoo! Finance or brokerages' websites. When seeking investment targets, retail investors must sort through the price history of a wide array of potential targets or all tradable stocks. Constantly pulling these data over hundreds or thousands of stocks is simply not practical without an API, especially considering that price information and return calculations must be updated frequently. Thus, a one-time, one-by-one manual download is not a feasible solution, which was described by a user in the Yahoo! Finance Help Community complaining about the lost functionality after the API shutdown: "I really don't see why they're even 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."

On May 16, 2017, Yahoo! Finance API was abruptly shut down, and access to the data were eliminated (Online Figure E1). Although Yahoo! offered no explanation for the shutdown, it was speculated that one reason was Verizon's acquisition of Yahoo!, Inc.[10] Notably, visitors to the Yahoo! Finance web page could still manually download the spreadsheet files to access historical data. However, this procedure is extremely tedious and impractical. Essentially, the API shutdown put an end to bulk download of free historical price data.

The abrupt shutdown caught users by surprise as evidenced by the outrage in a thread with more than 250 replies in the Yahoo! Finance Help Community. Even the forum administrator was caught off guard (Online Figure E2). Numerous reactions to the shutdown also emerged on other forums, such as Stack Overflow and Quora (Online Appendix D). In addition to venting frustration, these posts typically asked for alternatives to Yahoo! Finance API. The main alternative, Google Finance API, had been shut down five years earlier,[11] so users on forums such as Stack Overflow named several alternatives that provided similar functionalities, but the content was either costlier or less extensive. Given the drawbacks of alternative APIs at the time, users created and perfected new tools, such as yfinance,[12] following the shutdown. Overall, it took time for users to replace Yahoo! Finance API with alternatives given the development cycles for new tools, the learning curve, and switching costs (purchase alternatives, code adjustments, and trading strategy updates to suit different data formats and content from alternative APIs). In addition, some users were waiting for Yahoo! Finance API to be reinstated online.

Taken together, the sizable user base and the abrupt nature of the Yahoo! Finance API shutdown provide a plausibly exogenous natural intervention for us to examine and understand how tech-enabled data access (and the removal) affects the behavior of a large group of retail investors.

2.2. Active Retail Investors

As with all investors, heterogeneity exists among retail investors in terms of financial sophistication and behavioral biases (e.g., see Barber and Odean 2013 for a recent review). On one end of the spectrum, some retail investors prefer passive trading strategies, for example, buy and hold value stocks or copy others' trading ideas (Von Gaudecker 2015, Apesteguia et al. 2020) and demonstrate fewer behavioral biases (Dhar and Zhu 2006). On the other end, active day traders tend to treat trading like gambling and bet on stocks without solid fundamental analyses of their investment targets (Grinblatt and Keloharju 2009). Despite the wide variation in trading skills, sophisticated retail investors (e.g., those with security selection skills) are relatively rare (U.S. Securities and Exchange Commission 2012, Barber and Odean 2013). The majority are less-sophisticated retail investors (hereafter, active retail investors). This group is the primary focus of our study and of the finance and accounting literature in general (e.g., Barber and Odean 2002, Kumar 2009, Blankespoor et al. 2019, Da et al. 2021).

On average, active retail investors earn poor returns in the long run (Barber and Odean 2000, French 2008, Barber et al. 2009b). In their seminal work, Barber and Odean (2000) show that retail investors typically lose money compared with simply holding index funds, and the more they trade, the more they lose. Retail investors' collective losses in the stock market are non-negligible. Based on detailed data in Taiwan, the aggregate trading losses of individuals are equal to 2.2% of Taiwan's total gross domestic product (Barber et al. 2009b). Although retail investors can perform well in short windows (Kaniel et al. 2008), they are generally net buyers of stocks with weak future performance over long horizons (Grinblatt and Keloharju 2000).

Several reasons contribute to active retail investors' underperformance. First, most retail investors suffer from an informational disadvantage because they have less access to relevant information than professional investors. Even when information is fully accessible, they tend to have limited bandwidth to identify and process relevant information. For example, Blankespoor et al. (2019) show that, when both earnings and return information are available, retail investors choose to trade on past stock prices and ignore fundamental information, possibly because of high information processing costs to understand and interpret earnings information. When active retail investors pay scant attention to firm fundamentals, they enter the market with a key information deficit.

Many active retail investors also treat investment like gambling and as a form of entertainment. Some researchers find that lottery participation is a partial substitute for retail trading as larger jackpots in lotteries correspond to lower retail trading activities in the stock market (Dorn et al. 2015, Gao and Lin 2015). As with lotteries, active retail investors often search for cheap bets, hoping to achieve huge returns on some winning stocks. They have a taste for stocks with lottery-like payoffs (i.e., stocks with a low price, high volatility, and high skewness), which offer a very small chance of generating huge returns and a high probability of producing losses (Kumar 2009). A case in point is the recent saga of AMC and GameStop.[13]

To make matters worse, individuals are generally overconfident in their own skills (Moore and Healy 2008, Huffman et al. 2022). For example, most retail investors tend to believe they are better traders than their peers, and they may also have overly narrow predictions of future stock returns, overestimating their performance forecasting ability and underestimating the risk (Barber and Odean 2013). Past studies link such overconfidence with excessive trading and poor investment outcomes (Barber and Odean 2002, Biais et al. 2005, Grinblatt and Keloharju 2009).

In short, active retail investors suffer from various deficiencies, ranging from an informational disadvantage and gambling-like trading preferences to overconfidence. Whereas FinTech could alleviate retail investors' informational disadvantage, it may simultaneously amplify their other deficiencies. Thus, the net effect remains an open question.

2.3. Tech-Enabled Access to Financial Data and Active Retail Investors

Tech-enabled financial data can have both positive and negative effects. On the plus side, convenient access to an enlarged information set (or the same information set at a reduced acquisition cost) may improve trading informativeness. Yahoo! Finance API alleviated the informational disadvantages of retail investors by streamlining the speedy acquisition of historical price data and fundamental information on a large scale. Past studies show that tech-enabled information provision can benefit retail investors (Gao and Huang 2020, Farrell et al. 2022). For example, retail investors made more informative trades after the implementation of the electronic data gathering, analysis, and retrieval system by the SEC, which enabled free access to fundamental information (e.g., annual reports) (Gao and Huang 2020). Because Yahoo! Finance API also provided 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 more informed trades.

However, retail investors mostly ignore fundamental information and rely excessively on return information (Blankespoor et al. 2019). As such, the dominant trading strategies among active retail investors are built on past stock prices. Da et al. (2021, p. 175) find that retail investors "extrapolate from stocks' recent past returns, with more weight on more recent returns." In addition to gambling-like trading preferences, active retail investors armed with convenient access to financial data could design schemes to pick winning tickets by sifting through a sea of recent price data. The ability to request updated price data on demand through the API and at a large scale facilitates such trading strategies. As prices fluctuate, these strategies require constant rebalancing, leading to frequent trades. Given these flawed investment strategies (Zhang and Zhang 2015, Da et al. 2021), the abundant historical and real-time price data that were accessible through Yahoo! Finance API might not have improved the informativeness of each trade, but may have instead led retail investors to keep repeating the same mistakes.

More importantly, abundant data may amplify active retail investors' overconfidence, leading to excessive trades with even worse quality. On top of individuals' innate overconfidence (Svenson 1981), the inflow of a vast amount of data may exacerbate retail investors' illusions of knowledge, precision, and control. The illusion of knowledge, also known as the over-placement bias, is the belief that one is better than the median person because of more knowledge or superior skills (Barber and Odean 2013). When large amounts of data are readily available to retail investors through an API, they may be encouraged to engage in data mining, uncover spurious patterns from the data, and equate these patterns with knowledge. The greater the volume of data at hand, the more likely they are to have the illusion that they are more knowledgeable than the average investor in the market.

A related (but different) type of overconfidence is a false sense of accuracy and certainty. Formally, the illusion of precision (also known as over-precision or miscalibration) is the excessive certainty in the accuracy of one's judgment (Barber and Odean 2013, Moore 2023). Prior research shows that, when people are given more information to make a forecast, the perceived accuracy of their forecasts increases much faster than their actual accuracy (Oskamp 1965, Hall et al. 2007). Although API-enabled access to data and analyses might slightly improve retail investors' predictions of future price movements on the margin, convenient access to a long time series of data points for many stocks could also lead them to overestimate their improved accuracy and underestimate the uncertainty associated with future realization.

Finally, because an API can streamline the process of stock selection and portfolio monitoring, it induces an illusion of control—a misconception that one can influence the outcome of chance events (Langer 1975, Barber and Odean 2002). When retail investors play with a large amount of API-enabled financial data, they can identify investment targets from a wide range of stocks. Once a portfolio is formed, the API can also track real-time performance and help rebalance the portfolio if necessary. As personal involvement can elevate individuals' sense of control (Langer 1975), developing an investment strategy based on large-volume data may make an investor overly confident. Moreover, the mere access to an institutional-like data feed may already make some investors feel very much in control and conflate their control over actions (which investments to make) with control over outcomes (realized returns on the investments).

To sum up, although access to tech-enabled financial data has the potential to benefit retail investors by enlarging their information sets, it may strengthen their reliance on feedback trading, which automatically feeds historical prices and reinforces their tendency to ignore fundamental variables. More importantly, access to a large amount of financial data through APIs may disproportionately exacerbate active retail investors' overestimation of their knowledge, inflate their sense of forecasting accuracy, and induce a false sense of control in trading. The overconfidence (inflated by access to API-enabled data) is likely to lead to more trades and worse trading quality (Barber and Odean 2002, Statman et al. 2006).

3. Quasi-Natural Experiment

Our empirical design consists of two parts: a quasi-natural experiment and a randomized controlled experiment. The quasi-natural experiment leverages the Yahoo! Finance API shutdown to evaluate the effect of access to tech-enabled financial data on active retail investors' actions as a whole. The randomized controlled experiment directly tests individuals' behavioral biases when subjects are randomly assigned access to tech-enabled financial data. Together, the two experiments complement one another by evaluating the aggregate effect in the field and revealing the underlying individual-level mechanism in a laboratory setting. Before describing the randomized controlled experiment, we first lay out the identification strategy, sample and data, and empirical analyses for the quasi-natural experiment.

3.1. Identification

The abrupt shutdown of Yahoo! Finance API provides a quasi-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 have been Yahoo! Finance API consumers; therefore, they were disproportionately affected by the shutdown. As we do not possess investor-level data in the quasi-natural experiment, we infer active retail investors' trades from the stocks in which they prefer to invest. 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) because of their preferences or limitations (Barber and Odean 2000, Ivković et al. 2008, Grinblatt and Keloharju 2009, Conrad et al. 2014, Gao and Lin 2015). Therefore, we construct the control and treatment groups based on the stocks on which retail investors prefer to trade.

Empirically, we use various proxies to identify stocks favored by retail investors (hereafter, retail-favored stocks). First, following the spirit of past studies (Sias and Starks 1997, Kumar and Lee 2006, Gao and Lin 2015), we identify retail-favored stocks as those held primarily by retail investors. Their ownership structure represents a snapshot that uncovers the revealed preferences of retail investors. Retail holdings are measured as shares that are not reported to be held by institutional investors, scaled by total shares outstanding. We construct a dummy variable for retail-favored stocks (*RFS)*, which equals one when retail holdings are above the sample median to indicate retail-favored stocks (Table 1 documents all variable definitions). This provides a natural distribution of stocks ($RFS = 1$) whose retail trades were mostly impacted by the API and other stocks whose retail trading were less impacted (i.e., control firms; $RFS = 0$). In other words, retail-favored stocks make up the treatment group, whereas other stocks fall into the control group.

We also construct alternative proxies to identify stocks preferred by active retail investors. As many active retail

Table 1. Variable Definitions

Panel A. Variable definitions for the quasi-natural experiment	
Variables	Definitions
Retail_Vol	The volume 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).
Ab_Retail_Vol	*Retail_Vol* minus its median for the same day of the week over the past 10 weeks.
Ab_Retail_Buy	The volume of buy transactions initiated 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. Buy transactions are signed following Lee and Ready (1991).
Ab_Retail_Sell	The volume of buy transactions initiated 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. Buy transactions are signed following Lee and Ready (1991).
AIM	Amihud (2002) illiquidity measure, the natural logarithm of the ratio of absolute stock return to dollar volume $[1{,}000{,}000 \times \lvert ret \rvert \div (prc \times vol)]$
Spread	Daily bid–ask spread based on CRSP data, $100 \times (ask - bid)/[(ask + bid)/2]$.
CAR[1 W]	Cumulative abnormal buy-and-hold return for the next week, starting from the next trading day. Abnormal return is the delisting-adjusted individual stock return minus the corresponding market return.
CAR[iW, jW]	Cumulative abnormal buy-and-hold return from week *i* through week *j*, both inclusive ($i \neq 0$). Abnormal return is the delisting-adjusted individual stock return minus the corresponding market return.
Institutional_Vol	The volume of trades initiated by institutional investors, scaled by total shares outstanding and multiplied by 100. Institutional trades are nonretail trades with trade sizes above $50,000, following Bushee et al. (2020).
Ab_Institutional_Vol	*Institutional_Vol* minus its median for the same day of the week over the past 10 weeks.
RFS	Indicator for retail-favored stocks, equal to one if the stock's retail holding (measured as shares not reported to be held by institutional investors scaled by total shares outstanding) is above the sample median, zero otherwise.
Low_Priced	Indicator for low-price stocks, equal to one if the stock price is below the sample median, zero otherwise. The stock price is measured as of the last trading day before the main sample starts.
High_Volatility	Indicator for high-volatility stocks, equal to one if the idiosyncratic stock volatility is above the sample median, zero 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).
High_Skewness	Indicator for high-skewness stocks, equal to one if the stock skewness is above the sample median, zero 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).
Lottery_Like	Indicator for lottery-like stocks, equal to one if *Low_Priced* = *High_Volatility* = *High_Skewness* = 1, zero otherwise.
Small_Cap	Equal to one if the stock market capitalization is below the sample median, zero otherwise. We measure the market capitalization as of the last fiscal period end before the sample starts.
Post	Indicator for the post period, equal to one if the date is on or after May 16, 2017, zero otherwise.
Ret	Delist adjusted stock returns.
Ret^2	Square of *Ret*.
News	News coverage, measured as the natural logarithm of one plus the number of news articles on the Dow Jones Edition of RavenPack with a relevance score above 20 (the company name can be identified somewhere in the story).

Table 1. (Continued)

Panel A. Variable definitions for the quasi-natural experiment	
Variables	Definitions
Size	Firm size, measured as the natural logarithm of market capitalization at the fiscal period end.
BTM	Book-to-market ratio, measured as the ratio of the book value of the equity to its market value.
ROA	Return on assets (*ib/at*).
Loss	Indicator for loss, equal to one if *ROA* < 0, zero otherwise
R&D	R&D intensity (*xrd/at*).
Advertising	Advertising intensity (*xad/at*).
Leverage	Financial leverage ((*dltt* + *dlc*)/*at*).
Analysts	Financial analyst coverage, measured as the natural logarithm of one plus the number of financial analysts following the company.

Panel B. Variable definitions for the randomized controlled experiment	
Variables	Definitions
Amount	Trading intensity, measured as ($3,000 − cash remaining in the account)/$3,000, where $3,000 is the amount allocated to the account for every experiment subject.
Return	Trading performance, measured as the total dollar amount of gains/losses four weeks after the trade date, divided by $3,000.
Illusion of knowledge (pretest)	Belief that one is better than the median person because of more knowledge or superior skills, measured as negative one times the average of (1) subject-estimated percentage of other investors in the same brokerage house with superior investment skill and (2) similar percentage estimate on investment returns.
Illusion of knowledge (posttest)	Negative one times the average of (1) subject-estimated percentage of other investors in the trading game with superior investment skill and (2) similar percentage estimate on investment returns.
ΔIllusion of knowledge	Illusion of knowledge (posttest) minus Illusion of knowledge (pretest).
Illusion of precision (pretest)	Excessive certainty in the accuracy of one's judgment, measured as negative one times the average of (1) the difference between upper and lower bound estimations of the current unemployment rate, (2) the difference between upper and lower bound estimations of the unemployment rate one year later, (3) similar difference of estimations on the current inflation rate, and (4) similar difference of estimations on the inflation rate one year later.
Illusion of precision (posttest)	Negative one times the average of (1) difference between the upper and lower bound estimations of future price of two random stocks selected from the 100 stocks in the experiment and (2) similar difference of estimations on stock return percentages.
ΔIllusion of precision	Illusion of precision (posttest) minus Illusion of precision (pretest).
Illusion of control	Misconception that one can influence the outcome of chance events, measured as negative one times the willingness (1–5 Likert) to switch to passive investment means (e.g., index funds, robo-advisors).
Experiment_Treat	Equal to one if a subject has API-like data access in the experiment, zero otherwise.
Experiment_Placebo	Equal to one if a subject can see the top five best-performing stocks, zero otherwise.
Gender	Equal to one if female, zero if male.
Age	Age in years.
Marital status	Equal to one if single, two if married, three if divorced, four if widowed.
Children	The number of children.
Job category	Equal to one if retired, two if a housewife/-man, three if a student, four if blue collar, five if white collar, six if self-employed, seven if a civil servant, eight if other.
Education	Equal to one if high school degree, two if college degree, three if other degree.
Income	Equal to one if no income, two if up to $50K, three if between $50K and $100K, four if between $100K and $150K, five if between $150K and $200K, six if greater than $200K.
Investment experience	Equal to one if no investment experience, two if up to one year, three if between one and three years, four if between three and five years, five if between 5 and 10 years, six if between 10 and 15 years, seven if more than 15 years.
Daily trading time	Equal to one if the daily trading time is less than half an hour, two if between half and one hour, three if between one and two hours, four if between two and four hours, five if between four and six hours, six if between six and eight hours, seven if more than eight hours.
Daily trading amount	Equal to one if daily trading amount is up to $500, two if between $500 and $1,000, three if between $1,000 and $3,000, four if between $3,000 and $5,000, five if between $5,000 and $7,000, six if between $7,000 and $10,000, seven if more than $10,000.
Subjective financial literacy	Self-assessed financial literacy (1–7 Likert)
Objective financial literacy	The average score of four questions on compound interest, inflation, stocks, and bonds.

investors treat trading as gambling and bet on lottery-like stocks (Grinblatt and Keloharju 2009), we construct a lottery-likeness index (*Lottery_Like*), which equals one when the stock has above-median volatility (*High_Volatility*), above-median skewness (*High_Skewness*), and below-median price (*Low_Priced*), following Kumar (2009).[14] 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 of generating extreme positive returns. For completeness, we also use volatility, skewness, and price level as a single defining characteristic to identify lottery-like stocks (the results are consistent and reported in Online Appendix F).

In the same spirit as lottery-like stocks, we construct an additional proxy based on market capitalization (Barber and Odean 2000, Kumar and Lee 2006, Barber et al. 2009b, Gao and Lin 2015). We measure market capitalization as of the last trading day, a month before the API shutdown (when the main sample starts). We define a dummy variable (*small-cap*) that equals one for stocks with below-median market capitalization. Table 2, panel A, presents the correlations among the three proxies and the three subindices of lottery-like stocks. They positively correlate with each other as they all intend to capture stocks preferred by active retail investors.

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 with data available in Compustat and the Center for Research in Security Prices (CRSP) both before and after the shutdown of the 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 and two weeks around the shutdown) for supplemental analyses and 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).[15] The idea behind the classification is that retail trades are often executed off-exchange and offer a small price improvement (usually a fraction of a cent) relative to the national best bids and offers (Boehmer et al. 2021). Specifically, we classify retail sell (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 conservative and ensure accuracy, trades with prices at a round penny or near the half-penny (0.4–0.6 cents, inclusive) are not classified.[16] Admittedly, this classification omits some retail trades as not every retail trade is off-exchange or receives a price discount. Boehmer et al. (2021) validate this classification method and find it picks up a majority of overall retail trading activity and highly correlates with some proprietary retail trading data. 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 the normal level (the

Table 2. Retail-Favored, Lottery-Like, and Small-Cap Stocks

Panel A. Correlation between different proxies for 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 (# firms = 1,952)			RFS = 0 (# firms = 1,993)			
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 (# firms = 1,257)			Lottery_Like = 0 (# firms = 2,687)			
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 (# firms = 1,973)			Small_Cap = 0 (# firms = 1,971)			
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**

Notes. Panel A of this table presents the correlations between retail-favored, lottery-like (and its three sub-indices), 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 subperiods before and after the shutdown, respectively. See Table 1 for detailed variable definitions.

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

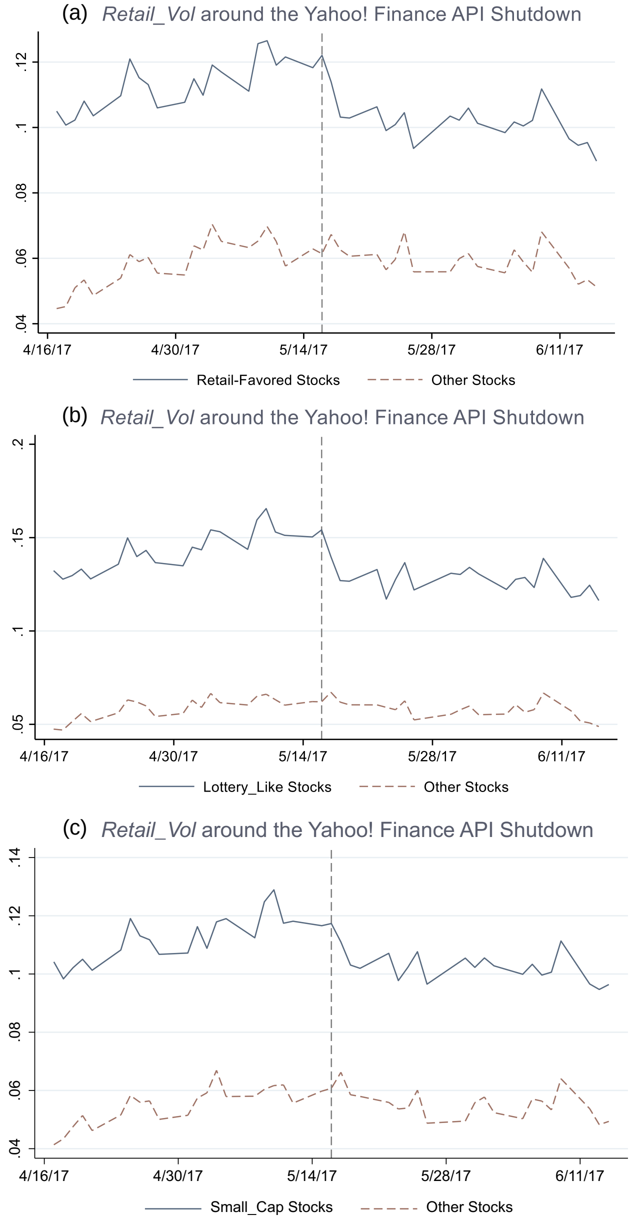
corresponding median scaled trading volume for the same day of the week over the prior 10 weeks) to construct the abnormal retail trading volume (*Ab_Retail_Vol*).

Table 2, panel B, splits the sample stocks by each treatment indicator and reports the change in retail trading volume around the API shutdown. *Pre* and *Post* indicate one month before and one month after the shutdown, respectively. The univariate analysis shows that both raw and abnormal retail trades fell significantly in the post period for retail-favored stocks. Specifically, *Retail_Vol* declines from 0.113 to 0.102 in retail-favored stocks (those with above-median retail holdings). This amounts to a 9.7% (= (0.102 − 0.113)/0.113) decrease in a month, which is statistically significant at 1%. Retail trading for other stocks stayed the same during the same period. Therefore, the univariate DID estimate (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 in retail-favored stocks decreased significantly 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 trades for retail-favored stocks, and the dashed line represents other stocks. Panels (a)–(c) in Figure 1 use *RFS*, *Lottery_Like*, and *Small_Cap*, respectively, to classify stocks favored by active retail investors. We observe a similar pattern across the three different classifications (Online Figure F1 exhibits similar plots for independent indices of lottery-like stocks). That is, 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 the DID analyses. Within three days after the shutdown, the solid line drops substantially, whereas the dashed line does not change noticeably except for a minor spike. The sharp decrease in retail trades immediately after the shutdown suggests that it is more likely to have occurred because of the shutdown than other confounding events.[17]

We obtain firm characteristics, stock performance, analyst following, news coverage, and institutional (retail) ownership from standard data sources (Compustat, CRSP, Institutional Brokers' Estimate System, RavenPack, and Thomas Reuters). To avoid look-ahead bias, firm characteristics are measured at the latest fiscal year ending before January 1, 2017. As it was several months before the API shutdown, these characteristics should have been available when trading decisions were made. We use firm characteristics as reported in annual reports (fourth quarter) rather than in quarterly reports to increase comparability across firms because annual reports are audited, and firms may display seasonality in different quarters.

Figure 1. (Color online) Retail Trading Volume Around the Shutdown of Yahoo! Finance API



Notes. This figure plots the daily retail trading volume (*Retail_Vol*) around the shutdown of Yahoo! Finance API for retail-favored stocks (solid line) and other stocks (dashed line). 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 and small-cap stocks as alternative proxies for retail-favored stocks. The vertical dashed lines indicate the shutdown of Yahoo! Finance API.

Table 3 reports the summary statistics for the main sample, which includes 169,430 daily observations from

Table 3. Summary Statistics

Panel A. Firm-day observations						
	N	Mean	Standard deviation	P25	P50	P75
Retail_Vol	169,430	0.083	0.200	0.010	0.025	0.065
Ab_Retail_Vol	169,430	0.019	0.131	−0.008	0.000	0.014
AIM	168,921	0.073	0.246	0.000	0.002	0.017
Spread	169,428	0.509	0.990	0.037	0.117	0.456
Institutional_Vol	169,430	0.081	0.168	0.000	0.020	0.083
Ab_Institutional_Vol	169,430	0.032	0.134	−0.009	0.000	0.025
Ret	169,430	0.000	0.023	−0.010	0.000	0.011
Ret^2	169,430	0.001	0.002	0.000	0.000	0.000
News	169,430	0.225	0.535	0.000	0.000	0.000

Panel B. Firm-level observations						
	N	Mean	Standard deviation	P25	P50	P75
Size	3,945	6.760	2.052	5.295	6.762	8.140
BTM	3,945	0.522	0.568	0.222	0.430	0.712
ROA	3,945	−0.072	0.291	−0.048	0.012	0.052
Loss	3,945	0.353	0.478	0.000	0.000	1.000
R&D	3,945	0.068	0.159	0.000	0.000	0.048
Advertising	3,945	0.009	0.025	0.000	0.000	0.003
Leverage	3,945	0.258	0.244	0.043	0.211	0.406
Analysts	3,945	0.967	0.930	0.000	0.693	1.609

Notes. 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 Table 1 for detailed variable definitions.

April 16, 2017, to June 15, 2017, for 3,945 unique stocks. To minimize the influence of outliers, we winsorize all variables at 1% and 99% except for dummy and log-transformed variables. On a typical day, retail trades 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 and is regularly covered by the media and financial analysts. The average ROA is negative, whereas the median ROA is slightly positive.

To measure the quality of retail trades, we study their predictive power for future returns following prior literature (Barber et al. 2009a; Kelley and Tetlock 2013, 2016; Gao and Lin 2015). Buy (sell) trades that positively (negatively) correlate with future stock returns are likely to be more profitable, on average, which reflects better investment decision making. 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 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 of the Quasi-Natural Experiment

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 DID regression specification:

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

where i represents the firm and t represents the date. The outcome variable is abnormal retail trades (*Ab_Retail_Vol*). The time indicator *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 the API shutdown.[18] W_{it} represents a set of firm day–level control variables, including the stock return, the square of the stock return, and news coverage (Da et al. 2011). Date fixed effects control for any changes in macroeconomic conditions that affect retail-favored stocks and other stocks. Firm fixed effects control for time-invariant firm heterogeneity. In the robustness checks, we also use lottery-like stocks and stocks with small market capitalization to identify the treatment group. That is, we replace *RFS* with *Lottery_ Like* and *Small_Cap*.

4.1.2. Empirical Results on Retail Trading Volume. We report the regression results of abnormal retail trading volumes in Table 4. Panel A uses retail investors' revealed preferences (*RFS* as determined by retail holdings), and 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, whereas columns (3) and (4) (columns (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 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 ($Post \times RFS$), which is significant at the 1% level. This result suggests that abnormal retail trades dropped significantly after the shutdown for retail-favored stocks, which is consistent with the idea that the API led active retail investors to trade more frequently.[19] Once we extend the sample period to four months, in columns (5) and (6), the coefficient on $Post \times RFS$ becomes smaller and statistically insignificant, suggesting that active retail investors gradually found alternative data sources to substitute for Yahoo! Finance API.

For control variables, we find that retail trading increases on days with larger stock movements and more news coverage, which is 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 and firms with intensive R&D, possibly because these firms are too complicated for retail investors to easily understand. Moreover, retail investors are more active in firms with intensive advertising expenditure, 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: $Ab\_Retail\_Vol_{it} = \alpha + \sum_{j=-4}^{4}\beta_j \cdot \text{Week}[j]_t \times RFS_i + \gamma \cdot W_{it} + \text{Date FE} + \text{Firm FE} + \epsilon_{it}$, where Week[$j$] equals one 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 and the corresponding confidence intervals in Figure 2. We find no significant difference in retail trading volumes for the control and treatment firms before the API shutdown, which further supports the parallel trend assumption.

Panel B of Table 4 reports the results using two alternative sets of proxies for retail-favored stocks: (1) stocks with lottery-like payoff structures and (2) stocks with smaller market capitalization. Specifically, columns (1)–(3) use *Lottery_Like*, a composite index of the three subcomponents (*Low_Priced*, *High_Volatility*, and *High_Skewness*) as the proxy, whereas columns (4)–(6) use *Small_Cap*. Across the two proxies, we observe a consistent pattern that abnormal retail trading volume decreases for retail-favored stocks immediately after the API shutdown and gradually climbs back about two months after the shutdown. Online Table F2 reports similar results using each of the subcomponents of the lottery-like stock index.

Our DID design alleviates concerns that macroeconomic shocks drive our results as these shocks would affect both retail-favored stocks and other stocks alike. Firm fixed effects control for firm-specific, time-invariant heterogeneity. It is still plausible, however, that some confounding events might affect retail-favored stocks; hence, active retail investors and other investors would reduce trading on them. To address this concern, we examine abnormal institutional trading volume as a placebo. We follow Bushee et al. (2020) and classify institutional trades from TAQ as nonretail trades larger than $50,000, which is consistent with the convention in the finance and accounting literature that assumes larger trades are likely initiated by institutional investors. Admittedly, this classification misses some institutional trades because institutional investors sometimes break big trades into smaller ones. We scale institutional trades and subtract the normal level to construct the abnormal institutional trading volume (*Ab_Institutional_Vol*) in the same fashion as the retail trades. If our main results on retail trades were caused by other events, we should observe a similar decrease in institutional trades, but that is not what we find. In Table 5, we present the results on abnormal institutional trading volume in a two-month window centered around the API shutdown (replacing *Ab_Retail_Vol* in the main analyses with *Ab_Institutional_Vol*). The coefficients on the interaction terms are insignificant across all proxies of retail-favored stocks. These nonresults give us more confidence in attributing our findings to the shutdown of the Yahoo! Finance API rather than to other confounding events.

There is still a concern that our results capture spurious trends. For example, retail investors are more prevalent in firms with high retail ownership, and they trade less frequently in June than in May because of summer vacations. To rule out these possibilities, we repeat our 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

Table 4. Retail Trades Around the Shutdown of Yahoo! Finance API

Panel A. Retail-favored stocks measured by holdings by retail investors

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable = *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)	
BTM	−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 fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	No	Yes	No	Yes	No
Firm fixed effects	No	Yes	No	Yes	No	Yes
Observations	43,374	43,374	169,430	169,430	326,675	326,675
R^2	0.230	0.542	0.211	0.362	0.213	0.291

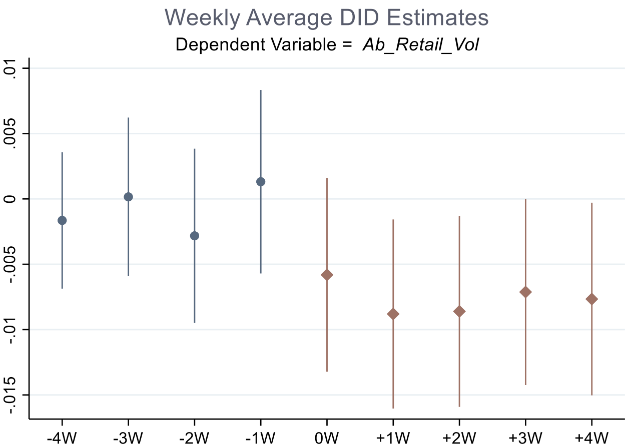
Panel B. Alternative proxies for retail-favored stocks

Dependent variable = *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 fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	39,433	169,426	326,635	39,433	169,426	326,627
R^2	0.558	0.362	0.291	0.557	0.362	0.291

Notes. 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-month, 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 Table 1 for detailed variable definitions. Robust standard errors clustered by firm are reported in parentheses.

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Figure 2. (Color online) Weekly DID Estimates



Notes. 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.

falsification sample. Table 6 shows no significant change in abnormal retail trading volume for retail-favored stocks after the pseudo shutdown in 2016. This nonresult reassures us that spurious trends do not drive our findings.

4.2. Retail Trade Quality

To test the quality of retail trades, 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 because of the inherent nature of their trading) for the subsequent cumulative abnormal return. The predictivity is estimated by the Fama–Macbeth regressions specified as[20]

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

where the dependent variable is future cumulative abnormal returns (CAR), that is, the buy-and-hold return of the individual stock minus the corresponding market return over different horizons. The key variable of interest is *Ab_Retail_Buy (sell)*, which measures abnormal buying (selling) volume by retail investors (scaled by outstanding shares). A more positive (negative) β_1 (β_2) indicates that overall retail buys (sells) have a higher predictivity for future returns and, hence, are more informative, on average. This is a common method in the finance literature to quantify the aggregated informativeness of a given group of investors.[21] Z represents a set of firm-day level control variables (see Table 7, panel A), following Kelley and Tetlock (2013).

Table 7 reports the regression results on the aggregated informativeness of retail buys and sells made in the two-month window around the shutdown.[22] To establish a benchmark and see how far ahead aggregated retail trades can predict future returns, we use the entire sample (retail-favored plus other stocks for

Table 5. Placebo Tests Using Institutional Trades

Dependent variable = *Ab_Institutional_Vol* Treat =	(1) *RFS*	(2) *Lottery_Like*	(3) *Small_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 fixed effects	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes
Observations	169,430	169,426	169,426
R^2	0.194	0.194	0.194

Notes. 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 4. 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 1 for detailed variable definitions. Robust standard errors clustered by firm are reported in parentheses.

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table 6. Falsification Tests

Dependent variable = *Ab_Retail_Vol* Treat =	(1) *RFS*	(2) *Lottery_Like*	(3) *Small_Cap*
Post × *Treat*	0.001	0.001	−0.000
	(0.002)	(0.003)	(0.002)
Ret	0.015	0.015	0.015
	(0.016)	(0.016)	(0.016)
Ret^2	20.374***	20.377***	20.376***
	(0.475)	(0.475)	(0.475)
News	0.020***	0.020***	0.020***
	(0.001)	(0.001)	(0.001)
Day fixed effects	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes
Observations	157,452	157,452	157,452
R^2	0.358	0.358	0.358

Notes. 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 4. 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 1 for detailed variable definitions. Robust standard errors clustered by firm are reported in parentheses.

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

the two-month window centered around the shutdown) and measure CAR (for each day) over the horizon of the subsequent week (*CAR[1 W]*), weeks 2–4 inclusive (*CAR[2 W, 4 W]*), and weeks 5–8 inclusive (*CAR[5 W, 8 W]*). The results across the six regressions using the entire sample (columns (1)–(6) in panel A) suggest that the significant results are concentrated in the first month (*CAR[1 W]* and *CAR[2 W, 4 W]*), and the significance fades as the horizon extends to the second month (*CAR[5 W, 8 W]*). Therefore, subsample analyses for the control and treatment groups are conducted for *CAR[1 W, 4 W]* (CAR for the subsequent month).[23]

Next, we run Fama–Macbeth regressions separately for retail-favored stocks and other stocks in both the preperiods and postperiods and perform a DID 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

Table 7. The Predictivity of Retail Trades for Future Returns

Panel A. Return predictability of retail trades						
Dependent variable = Sample	(1) *CAR[1 W]* & All	(2)	(3) *CAR[2 W,4 W]* & All	(4)	(5) *CAR[5 W,8 W]* & All	(6)
	Pre	Post	Pre	Post	Pre	Post
Ab_Retail_Buy	0.308	1.586**	−0.491	2.870*	−0.274	1.923
	(0.704)	(0.657)	(1.452)	(1.515)	(1.760)	(1.497)
Ab_Retail_Sell	−2.287***	−2.950***	−1.267	−2.456**	−3.322	2.183
	(0.679)	(0.691)	(1.282)	(1.083)	(2.726)	(1.398)
Ret^2	−109.587***	−120.814***	−97.188**	−24.747	−62.483	−215.050**
	(30.127)	(28.208)	(39.607)	(56.798)	(91.395)	(77.352)
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	0.069*	0.063	0.283***	−0.053	0.027	0.364**
	(0.035)	(0.042)	(0.036)	(0.081)	(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	3.964	−6.461**	3.372	−5.103	11.962*	0.705
	(2.853)	(2.766)	(2.925)	(4.404)	(6.296)	(2.917)
CAR[-1 W]	0.018	−0.017	0.057**	−0.034***	0.052	0.020
	(0.016)	(0.014)	(0.024)	(0.012)	(0.049)	(0.016)

Table 7. (Continued)

Panel A. Return predictability of retail trades

Dependent variable = Sample	(1)	(2)	(3)	(4)	(5)	(6)
	CAR[1 W] & All		*CAR[2 W,4 W]* & All		*CAR[5 W,8 W]* & All	
	Pre	Post	Pre	Post	Pre	Post
CAR[-2 W,-4 W]	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^2	0.017	0.026	0.022	0.014	0.018	0.024

Panel B. Return predictability of retail trades in RFS and other stocks

Dependent variable = Sample	(1)	(2)	(3)	(4)
	CAR[1 W,4 W] RFS = 0		*CAR[1 W,4 W]* RFS = 1	
	Pre	Post	Pre	Post
Ab_Retail_Buy	7.570**	−0.950	−1.012	7.316***
	(3.504)	(2.543)	(1.626)	(1.384)
Ab_Retail_Sell	−2.372	−3.961	−5.210***	−6.376***
	(3.290)	(3.822)	(1.142)	(1.191)
Ab_Retail_Buy (column (4) − column (3)) – (column (2) − column (1)):		16.848***		
Ab_Retail_Sell (column (4) − column (3)) – (column (2) − column (1)):		0.423		
Controls	Yes	Yes	Yes	Yes
Observations	41,837	43,645	40,949	42,705
R^2	0.032	0.033	0.020	0.015

Panel C. Return predictability of retail trades in lottery-like and other stocks

Dependent variable = Sample	(1)	(2)	(3)	(4)
	CAR[1 W,4 W] Lottery_Like = 0		*CAR[1 W,4 W]* Lottery_Like = 1	
	Pre	Post	Pre	Post
Ab_Retail_Buy	5.094*	1.377	−1.448	7.777***
	(2.918)	(2.419)	(2.024)	(1.145)
Ab_Retail_Sell	−4.623**	−8.168***	−4.345***	−4.843***
	(2.186)	(1.906)	(1.229)	(1.501)
Ab_Retail_Buy (column (4) − column (3)) – (column (2) − column (1)):		12.942***		
Ab_Retail_Sell (column (4) − column (3)) – (column (2) − column (1)):		3.047		
Controls	Yes	Yes	Yes	Yes
Observations	56,418	58,857	26,368	27,489
R^2	0.020	0.031	0.024	0.017

Panel D. Return predictability of retail trades in small-cap and other stocks

Dependent variable = Sample	*CAR[1 W,4 W]* Small_Cap = 0		*CAR[1 W,4 W]* 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 (column (4) − column (3)) – (column (2) − column (1)):		9.668**		
Ab_Retail_Sell (column (4) − column (3)) – (column (2) − column (1)):		−2.319		
Controls	Yes	Yes	Yes	Yes
Observations	41,387	43,231	41,399	43,115
R^2	0.038	0.047	0.016	0.020

Notes. 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 two-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) and (2), from week 2 to week 4 in columns (3) and (4), and from week 5 to week 8 in columns (5) and (6). The key variable of interest is abnormal retail buys and sells (*Ab_Retail_Buy*, *Ab_Retail_Sell*). Panel B reports separate regressions in retail-favored stocks (*RFS*) and other stocks. Panel C (D) reports separate analyses for lottery-like (small-cap) stocks and other stocks. See Table 1 for detailed variable definitions. Standard errors with lags of two are reported in parentheses (Newey and West 1987).

***, **, and * indicate statistical significance at the 1%, 5%, and 10%, respectively.

Table 8. Market Liquidity

Panel A. AIM			
Dependent variable = *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 fixed effects	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes
Observations	168,921	168,917	168,917
R^2	0.612	0.612	0.612
Panel B. Spread			
Dependent variable = *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 fixed effects	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes
Observations	169,428	169,424	169,424
R^2	0.753	0.753	0.753

Notes. 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 Table 1 for detailed variable definitions. Robust standard errors clustered by firm are reported in parentheses.

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

B (*RFS* 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 find no significant changes for retail sells, possibly because selling stocks could be driven by liquidity, independent of the availability of API-enabled historical data. This evidence suggests that 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. This finding is consistent with changes in investors' overconfidence as the underlying mechanism.

4.3. Market Liquidity

As improved market liquidity is often a sign of uninformed trades, we study the change in market liquidity around the shutdown of Yahoo! Finance API. Past research shows that uninformed trading in the stock market improves market liquidity (Grossman and Miller 1988, Peress and Schmidt 2020). More uninformed trading makes it easier for a given investor to find a counterparty. More importantly, in a market with relatively more uninformed trades, investors are less concerned with adverse selection (trading with counterparties with an information advantage). Consequently, investors are more willing to trade with each other, resulting in a lower price impact per share traded and a lower bid–ask spread (Greene and Smart 1999, Han et al. 2016). Based on this intuition, if historical price data induce more uninformed retail trading, we should see better liquidity before the shutdown (i.e., liquidity deteriorates after the API shutdown).

Table 8 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.

5. Evidence of the Underlying Mechanism: Randomized Controlled Experiment

5.1. Experimental Design

We conduct a randomized controlled experiment to empirically test the underlying mechanism through which tech-enabled data access influences individuals' trading behavior. The experiment exposes a group of participants to tools that are similar to those available through Yahoo! Finance web pages with or without APIs. Without the API, Yahoo's Finance page provides two types of information. First, it allows individuals to search for a trading target (by referencing the name of the target company or its stock symbol) and to gather information about that target's financial statistics and historical data. Second, it displays information about the best performing stocks, which facilitates investors' search for investment targets. With API, investors can bulk download financial statistics and historical data for all targets covered in Yahoo! Finance's databases, thereby automating investment target screening and portfolio monitoring.

We design our experiment to mimic these three conditions in which retail investors use Yahoo! Finance. We randomly assign subjects to one of the following experimental conditions: (a) single-company search tool (all company data: statistics and historical data) (control condition), (b) single-company search tool + 5 best-performing companies based on their 52-week return (placebo condition), and (c) single-company search tool + 5 best performing companies based on their 52-week return + spreadsheet access to all companies' historical price data (treated condition).[24] The single-company search function available to all experimental groups mimics the search function available on the Yahoo! Finance web page before and after the API shutdown. The placebo condition is created to proxy for providing plain information (not large-scale data). To minimize the technical hurdle for the randomly assigned participants to use the API data, we mimic the actual API function by creating a button that allows participants to download a flat CSV file that contains historical data of stock targets in our experiment.[25] Online Appendix J presents details regarding the execution of the experiment (e.g., screenshots of the landing pages for each experimental group).

To manage the time needed to complete the trading task, we give all participants 100 real publicly traded stocks as potential investment targets.[26] The stocks' identifiable information (e.g., ticker, company name) is hidden and replaced with pseudonyms (Company 1 to Company 100) to avoid contamination by prior knowledge or information. Each participant is given 20 minutes to select investment targets, a fixed $3,000 budget, and 10 minutes to assign the budget to their desired targets with the goal of maximizing investment performance within a one-month investment horizon. The unallocated budget is treated as cash investments.

To increase the external validity of the experiment, we put design features in place that select participants with some day-trading experience. First, we recruit the subjects from major trading subReddits—r/investing, r/stocks, r/algotrading, r/fintech, and r/finnhubAPI—by sharing the experiment links with active participants in those threads as well as by encouraging those contacted to share the link with other day-traders they know. Next, we design the first round of our experiment to assess participants' financial literacy (see Online Appendix K for survey questions). To ensure that only those with sufficient financial literacy take part in our experiments,[27] we eliminate and compensate ($7) those who cannot answer half or more of the literacy questions. Round two of the experiment is three days after the first round to avoid priming the subjects. Subjects surviving the first round compete to create investment portfolios. In this round, subjects are assigned to the three experimental conditions mentioned earlier. Subjects in the second round are paid $7 for simply completing the task and an additional $20 if their return performance (based on the dollar value of the return on the portfolio of investment 30 days after the investment) is ranked in the top 10%[28] and $10 if their return performance is ranked between the top 11% and 20%. The monetary incentives ensure that participants take their investments seriously. We estimate returns on portfolios based on the real return performance of the 100 targets represented in the experiment. The investment targets' financial statistics, historical data, and the page on the five best performing stocks in the past 52 weeks are all based on real information extracted from Yahoo! Finance.

5.2. Measurements in the Experiment

Based on the experimental design, the treatment variable, *Experiment_Treat*, equals one if a person has API-like data access (group C) and zero otherwise. The *Experiment_Placebo* indicator equals one if a person does not have data access but can see the top five best performing stocks (group B). A subject is placed in the control group (group A) if both indicators equal zero. Hence, the two indicators capture the three subject groups as described in the experimental design.

To account for known confounding factors, we control for several individual characteristics. Because males typically trade more and lose more than females (Barber and Odean 2001), we ask subjects to report their gender in the survey (Online Appendix K). As individuals' life stage, family situation, and financial status may affect their risk tolerance, the survey records participants' age, marital status, number of children, job category, education, and income (Dorn and Huberman 2005). We also control for individuals' trading experience (number of years) and trading habits (daily trading time and amount). We measure financial literacy in two ways: subjectively and objectively. For subjective financial literacy, we ask individuals to self-assess their financial knowledge on a scale from one to seven. For objective financial literacy, we test subjects on their basic financial knowledge in compound interest, inflation, stocks, and bonds (Bursztyn et al. 2014).

In terms of outcomes, we first assess trading amount and profitability in the randomized controlled experiment. We measure investment amount as ($3,000 − cash remaining)/$3,000, which is a percentage of the invested amount out of all available cash for each subject. Individuals' return is measured as the total dollar amount of gains/losses four weeks after the trade date and scaled by $3,000. The two measures are counterparts for retail trading volumes and return predictability in the quasi-natural experiment.

We next measure each subject's overconfidence as reflected in their illusions of knowledge, precision, and control. In the pretest (before the trading game), the illusion of knowledge is measured as subjects' estimations of their investment skills and returns relative to other investors in the same brokerage house (estimations of the percentage of people with higher capability/returns). In the posttest (after the trading game), illusion of knowledge is measured as subjects' estimations of their skills and investment returns relative to other participants in the trading game. The survey questions are adapted from Glaser and Weber (2007). The pretest level of the illusion of knowledge allows us to set a baseline before participants start the trading game, and we construct the change in the illusion of knowledge, Δ*Illusion of knowledge*, as the posttest score minus the pretest score.

In the pretest, we measure the illusion of precision as the difference between subjects' upper and lower bound estimations of current and future macroeconomic indicators (unemployment rate, inflation rate) (Ortoleva and Snowberg 2015). In the posttest, illusion of precision is measured as the difference between the upper and lower bound estimations of future price and returns of two random stocks selected from the 100 stocks in the experiment. The narrower the interval, the higher the overconfidence. Δ*Illusion of precision* is again calculated as the difference in posttest and pretest scores.

The illusion of control is measured as the willingness to switch to passive investment options (e.g., index funds, robo-advisors). Because the sense of control is related to investments, this variable is measured only in the posttest (after investments have taken place). Details of the survey measurements are listed in Online Appendix K. For ease of interpretation, all overconfidence measures are multiplied by −1. Hence, when using any of the three variables as an outcome in a regression, a positive coefficient indicates an increase in overconfidence.

5.3. Findings of the Experiment

Before testing the effect of API-like data access, we ensure that the randomized assignments result in similar characteristics between subjects in the treatment group versus those in the control and placebo groups. The final sample of qualified subjects consists of 282 individuals with an average age between 27 and 28, 154 of whom are men. The subjects report significant trading experience (three years or more of experience in day trading) with all reporting at least 30 minutes per day, on average, of day trading in a usual week and most reporting less than $500 for daily trading (buying/selling). We randomly assign 100 participants to the treatment group, 92 participants to the placebo group, and the remaining 90 participants to the control group. Columns (1)–(3) in Table 9 present the mean of individual characteristics for the three groups. Column (4) reveals some differences in gender, education, and the number of children for subjects in the placebo group and treatment group. Column (5) shows that subjects in the treatment group have a slightly higher average income than those in the control group. Across all three groups, 95% of subjects report their gross annual income to be lower than $50,000, and the remaining 5% report no income. Thus, overall variation in income is extremely low. All other individual characteristics do not exhibit significant differences. Hence, the randomization process results in largely comparable individuals in the treatment group versus the control and placebo groups. Additionally, we ensure the

Table 9. Individual Characteristics (Randomized Experiment)

	(1) Control	(2) Placebo	(3) Treatment	(4) Placebo versus treatment	(5) Control vs. treatment
Gender	1.667	1.402	1.570	−0.168**	0.097
Age	26.689	27.478	27.400	0.078	−0.711
Marital status	1.389	1.446	1.450	−0.004	−0.061
Children	0.767	0.620	0.790	−0.170*	−0.023
Job category	3.756	3.815	3.830	−0.015	−0.074
Education	2.300	2.359	2.220	0.139**	0.080
Income	1.922	1.935	1.980	−0.045	−0.058*
Investment experience	4.522	4.391	4.500	−0.109	0.022
Daily trading time	3.033	3.098	3.030	0.068	0.003
Daily trading amount	1.011	1.033	1.020	0.013	−0.009
Subjective financial literacy	4.767	4.870	4.720	0.150	0.047
Objective financial literacy	1.850	1.837	1.810	0.027	0.040
Observations	90	92	100		

Notes. Columns (1)–(3) of this table present the means of individual characteristics in the control, placebo, and treatment groups. Columns (4) and (5) report the mean comparisons of the treatment group with the two other groups. All groups (including the control group) had access to the single-company search tool. The placebo group had additional access to the top five best performing stocks list. The treatment group had additional access to the best performing stocks list plus the master file containing all stocks' historical price data (treatment condition). See Online Appendix J for experiment executions and Online Appendix K for survey questions.

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table 10. Investment Amount and Return (Randomized Experiment)

Dependent variable	(1) *Amount*	(2) *Return*
Experiment_Treat	0.058***	−0.037***
	(0.008)	(0.006)
Experiment_Placebo	0.003	0.002
	(0.008)	(0.006)
Gender	0.000	−0.006
	(0.006)	(0.005)
Age	−0.001	−0.001
	(0.001)	(0.001)
Marital status	−0.001	−0.002
	(0.006)	(0.005)
Children	0.001	0.000
	(0.005)	(0.004)
Job category	−0.005	−0.005
	(0.007)	(0.005)
Education	0.000	0.009*
	(0.007)	(0.005)
Income	0.003	0.008
	(0.014)	(0.011)
Investment experience	−0.004	0.000
	(0.006)	(0.005)
Daily trading time	0.000	−0.001
	(0.003)	(0.003)
Daily trading amount	0.001	0.017
	(0.022)	(0.017)
Subjective financial literacy	0.000	0.003
	(0.003)	(0.002)
Objective financial literacy	−0.012	0.015
	(0.017)	(0.013)
Observations	282	282
R^2	0.235	0.231

Notes. This table reports the regression results of the investment amount and return based on the controlled experiment. The dependent variable *Amount* is the amount of investment in the stocks (as a percentage of the $3,000 budget). *Return* is the total dollar amount of gains/losses four weeks after the trade date and scaled by $3,000. The key explanatory variable is *Experiment_Treat*, which equals one if a subject is in the treatment group (with access to the master file containing all stocks' historical price data) and zero otherwise. Treated subjects could also access the single-company search tool and the best performing stock list. *Experiment_Placebo* is an indicator for the placebo group (with access to the single-company search tool and the best performing stock list but not the master data file). See Table 1 for detailed variable definitions, Online Appendix J for experiment executions, and Online Appendix K for survey questions.

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

representativeness of our sample by comparing the demographic information of the subjects with known statistics and in related studies.[29]

After confirming the sample representativeness and the validity of the random assignment, we first test how access to API-like data access affects retail investors' investment outcomes. As in the quasi-natural experiment, we are interested in knowing the volume and quality of retail trades. Because the randomized experiment contains individual-level observations, we can observe how much money each person puts into the stock market. Column (1) in Table 10 shows that treated subjects invest more in the stock market, which is consistent with the retail trading volume findings in the quasi-natural experiment (Table 4). Column (2) shows that treated subjects earn a significantly lower return, which is consistent with the return-predictability findings (Table 7). We observe no significant change in trading amount or return for subjects in the placebo group, and the absolute values of coefficients on *Experiment_Treat* are significantly higher than those of *Experiment_Placebo* in both regressions. For the individual-level observations, the results demonstrate that tech-enabled data access induces individuals to trade more and lose more in the stock market, which provides additional support for our findings from the quasi-natural experiment.

More importantly, the individual-level observations in the experiment allow us to test retail investors' behavioral bias. We study whether a person becomes more overconfident when exposed to API-like data access. Columns (1) and (2) in Table 11 show that the increments in illusion of knowledge and illusion of precision are significantly higher for treated subjects. Column (3) demonstrates that the illusion of control is also higher for treated subjects. Across the board, treated subjects show higher levels of overconfidence as a result of their access to tech-enabled data. In contrast, placebo subjects do not experience changes in their overconfidence level. Consistent with prior findings (Barber and Odean 2002, Statman et al. 2006), overconfidence observed in this experiment leads to more and worse trades (Online Appendix L).

In sum, the quasi-natural experiment demonstrates the market-level volume and quality changes of retail trades induced by tech-enabled access to financial data. The randomized controlled experiment reveals the mechanism for the observed changes. We find that tech-enabled financial data access increases retail investors' overconfidence, leading to excessive trading and losses in the stock market.

6. Conclusions

Taken together, the quasi-natural and randomized controlled experiments paint a coherent picture of the impact of access to tech-enabled data on retail investors. The two experiments complement each other with the quasi-natural experiment documenting the realized effect using large-scale, stock-level analysis and the randomized controlled experiment delineating the underlying mechanism at the individual level. The quasi-natural experiment leverages the fact that (1) active retail investors were disproportionately affected by the Yahoo! Finance API shutdown as they relied more on real-time and historical price data that were accessible through the API and (2) these investors favored different types of stocks than other investors

Table 11. Overconfidence Induced by Tech-Enabled Data Access (Randomized Experiment)

Dependent variable	(1) *ΔIllusion of knowledge*	(2) *ΔIllusion of precision*	(3) *Illusion of control*
Experiment_Treat	0.199***	0.637***	1.181***
	(0.061)	(0.186)	(0.122)
Experiment_Placebo	−0.092	0.102	0.010
	(0.064)	(0.195)	(0.127)
Gender	−0.073	−0.145	0.012
	(0.052)	(0.157)	(0.103)
Age	0.006	0.027	−0.003
	(0.007)	(0.023)	(0.015)
Marital status	−0.009	−0.190	0.002
	(0.049)	(0.149)	(0.098)
Children	−0.008	0.200	−0.009
	(0.044)	(0.133)	(0.087)
Job category	−0.010	0.293*	0.059
	(0.053)	(0.162)	(0.106)
Education	−0.054	−0.086	−0.080
	(0.055)	(0.168)	(0.110)
Income	−0.089	0.063	−0.010
	(0.117)	(0.355)	(0.232)
Investment experience	−0.019	−0.055	0.021
	(0.050)	(0.152)	(0.099)
Daily trading time	−0.047*	0.056	−0.042
	(0.027)	(0.083)	(0.054)
Daily trading amount	−0.140	−0.229	0.011
	(0.179)	(0.543)	(0.355)
Subjective financial literacy	−0.023	0.122	0.013
	(0.025)	(0.076)	(0.050)
Objective financial literacy	−0.024	−0.318	−0.715***
	(0.134)	(0.408)	(0.267)
Observations	282	282	282
R^2	0.114	0.100	0.358

Notes. This table reports the regression results of overconfidence based on the randomized controlled experiment. Overconfidence is assessed in three dimensions: illusions of knowledge, precision, and control. Illusion of knowledge is the belief that one is better than the median person because of more knowledge or superior skills measured as subjects' estimations of their investment skills and returns relative to other investors. Illusion of precision is the excessive certainty in the accuracy of one's judgment measured as the difference between subjects' upper and lower bound estimations of the unemployment rate, inflation rate, and future price and returns of two randomly selected stocks. ΔIllusions of knowledge and ΔIllusion of precision are calculated as the posttest score minus the pretest score. Illusion of control is the misconception that one can influence the outcome of chance events, measured as the willingness to switch to passive investment options (only measured after the trading game). For ease of result interpretation, the survey scores on illusions of knowledge, precision, and control are multiplied by −1 so that higher values indicate higher overconfidence. The key explanatory variable is *Experiment_Treat*, which equals one if a subject is in the treatment group (with access to the master file containing all stocks' historical price data) and zero otherwise. Treated subjects could also access the single-company search tool and the best performing stock list. *Experiment_Placebo* is an indicator for the placebo group (with access to the single-company search tool and the best performing stock list but not the master data file). See Table 1 for detailed variable definitions, Online Appendix J for experiment executions, and Online Appendix K for survey questions.

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

(clientele effect). By comparing the retail trades in retail-favored stocks to other stocks, we find that accessing institutional-like financial data induces excessive retail trading with lower returns. Our randomized controlled experiment shows that, despite the potential to ease information frictions, convenient access to a large amount of raw historical price data exacerbates Main Street investors' illusions of knowledge, precision, and control, which induces more trades and leads to greater losses. With the caveat that retail investors who use APIs may not be generalizable to all retail investors, our findings converge with the theoretical predictions in Zhang and Zhang (2015, p. 19) that "more aggressive feedback trading creates higher risks for uninformed traders without bringing higher benefits."

Our study contributes to the literature on financial technologies and retail investors. Technology affects retail investment in two fundamental ways: reducing transaction costs and reducing information acquisition costs. In terms of transaction cost reduction, switching to online trading and zero-cost trading via mobile applications is shown to reduce trading friction for retail investors, leading to more frequent and worse trades (Barber and Odean 2002, Ozik et al. 2021). We focus on tech-enabled information acquisition cost reduction, which is conceptually different from transaction-cost reduction as greater access to information holds the promise of improving investors' decision-making quality. Additionally, whereas some studies point out that retail investors suffer from overconfidence (e.g., Barber and Odean 2001), our randomized controlled experiment allows us to directly measure retail investors' overconfidence and link it to access to the underlying technology.

In terms of tech-enabled information dissemination, whereas past studies study providing retail-grade information through technologies, we focus on the effect of providing institutional-like data to retail investors. Unlike information that is periodically updated (e.g., earnings announcements), stock price data are constantly updated in real time and are observed in much larger volumes. Financial data APIs make it possible for Main Street investors to access such high-volume, real-time, and historical price data in a scalable manner. For example, real-time and scalable delivery of financial data through APIs allows for a nimble and automated trading strategy. Thus, this study examines a fundamentally different type of information provision in terms of information content, the scale of transmission, and the technology that enables it. Prior studies show that online dissemination of accounting information and investment research helps retail investors make more informative trades (Gao and Huang 2020, Farrell et al. 2022). We complement and extend prior research by examining a new type of tech-enabled information dissemination and find that unlocking access to API-enabled, high-volume price data for Main Street investors can hurt them by inducing excessive trading.[30] Thus, our findings highlight the importance of scrutinizing not only the quality of information (Clarke et al. 2020, Ammann and Schaub 2021), but also the scale and type of information presented to retail investors.

Although providing more information can and should be useful to investors, our study points to a potential pitfall: it can create an illusory sense of knowledge, precision, and control for retail investors, which, in turn, leads to excessive trading and lower returns. As retail investors tend to trade on stock prices and ignore fundamental information (Zhang and Zhang 2015, Blankespoor et al. 2019, Da et al. 2021), technologies that democratize access to raw price data can inadvertently reinforce such preferences and further outweigh past returns relative to firm fundamentals in their trading strategies. When tech-enabled access to data amplifies retail investors' behavioral deficiencies, the investors become more vulnerable to their suboptimal trading strategies. Given the increasing penetration of financial technologies, combined with prior evidence that investor sophistication mitigates behavioral biases (Feng and Seasholes 2005, Dhar and Zhu 2006), our findings suggest that the SEC's call to improve financial education (Stein 2018) is even more urgent. Moreover, as psychological traits explain financial behaviors better than financial literacy (Fernandes et al. 2014), financial education should incorporate training to reduce behavioral biases and tendencies (e.g., overconfidence and overtrading), overreliance on price information when ignoring fundamentals, gambling-like behaviors, under-diversification, and obsession with winner picking. When Main Street investors' awareness of such biases increases through education, they can take appropriate action to mitigate or correct their biases, thus improving the overall quality of their trades.

Regulatory bodies, such as the SEC, have long been interested in democratizing access to financial technologies (e.g., https://www.sec.gov/edgar/sec-api-documentation) and monitoring and alerting the public about financial technologies and their risks.[31] Our findings highlight the risks of exposing retail investors to financial technologies that provide convenient access to voluminous and dynamic price history data, which retail investors frequently seek. As such, our study lays out areas of regulatory focus for governing bodies such as the SEC and provides information about the specific types of risks and biases that could be introduced when such unregulated technologies are made available. Whereas regulators have a wide range of tools to combat these risks, even some simple, passive forms of informing the public about the risks of consuming certain products are shown to be effective in curbing harmful consumption in other contexts (Hammond et al. 2006). Thus, this practice could also be easily implemented for financial technologies.

Acknowledgments

The authors thank Gregory Eaton, Huseyin Gulen, Keongtae Kim, Soyoon Kong, Alvin Leung, Ben Liu, and Min-Seok Pang as well as seminar and conference participants at the City University of Hong Kong, Santa Clara University, California State University Long Beach, 2021 National Bureau of Economic Research (NBER) Summer Institute (IT and Digitization), 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.

Endnotes

[1] See 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 and https://www.vox.com/the-goods/22249458/gamestop-stock-wallstreetbets-reddit-citron.

[2] Existing financial technologies broadly range from platforms enabling rapid diffusion of knowledge and information about trading, such as Seeking Alpha, to applications facilitating low (transaction) cost trading for households, for example, Robinhood. Online Appendix A provides a literature review of tech-enabled information provision for retail investors.

[3] APIs provide access to raw data at the back end of developed applications, such as Google Maps and Yahoo's financial data. In particular, data APIs utilize standardized data formats (e.g., JSON, CSV) and easy data transfer modes (e.g., generic HTTP POST/GET requests) to give access to often complicated back-end databases. APIs, therefore, are a critical component in enabling third parties to develop applications on an existing platform and democratize access to large and frequently updated data sets that are otherwise difficult to comb through.

[4] Illusions of knowledge, precision, and control are the primary, sometimes conflated, categories of overconfidence studied in the literature (Langer 1975, Moore and Healy 2008, Barber and Odean 2013, Ortoleva and Snowberg 2015). The illusion of knowledge is often known as overplacement or better-than-average bias. The illusion of precision is often known as over-precision or miscalibration.

[5] We randomly assign API-like data access in a randomized controlled experiment (discussed in detail in Section 5) and uncover findings consistent with that of the quasi-natural experiment, which also addresses the selection concern.

[6] See https://www.bloomberg.com/professional/product/server-api/.

[7] This API (http://ichart.finance.yahoo.com) was used by the majority of the Yahoo! Finance API community; the outcry immediately after it shut down confirmed its popularity. Some motivated users found a partial workaround combining an alternative API (https://query1.finance.yahoo.com/) with tokens. However, this method has several drawbacks: it requires cookies and has query limits, it is less reliable with unexpected missing values, prices are computed differently (adjusting for dividends or not), and the date is in a different format. These drawbacks discouraged switching. Even if users switch, there is a learning curve, and some adjustments are needed, precluding a fast switch. Therefore, the shutdown in May 2017 still functionally removed the API tool for the majority of Yahoo API users.

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

[9] See https://www.amibroker.com/.

[10] See https://www.marketcast.co/blog.html.

[11] See https://groups.google.com/g/google-finance-apis/c/q-DbjbzQDGQ.

[12] yfinance is an open-source library providing access to market data on Yahoo! Finance. See https://pypi.org/project/fix-yahoo-finance/0.0.13/#history and https://algotrading101.com/learn/yfinance-guide/.

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

[14] Following Kumar (2009), volatility is defined as the standard deviation of the residual return from estimating a four-factor model on daily returns during six 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.

[15] Several recent accounting and finance studies follow Boehmer et al. (2021) to identify retail trades (Blankespoor et al. 2018, Bonsall et al. 2020, Bushee et al. 2020, Huang et al. 2021, Farrell et al. 2022).

[16] For example, prices such as \$100.003 and \$99.997 are classified as a retail sell and a retail buy, respectively. If a price is \$100.005 or \$99.995, it is unclassified.

[17] 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 paragraphs, we found no significant events or incidents that would have significantly affected retail investors.

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

[19] Online Table G1 shows similar patterns in trading volumes when retail buys and retail sells are separated. Online Table G2 tests a similar specification in which firms with retail holdings in the middle 20th percentile are removed. Online Table G3 tests the main specification using a matched control and treatment group on fundamental variables. The results in Online Tables G1–G3 are consistent with that of Table 4. Online Appendix H shows that retail investors utilized return rather than earnings information, which is consistent with anecdotes in Online Appendices D and E and prior research (Blankespoor et al. 2019, Da et al. 2021).

[20] For Fama–Macbeth regressions, we first run the regression separately in each period (i.e., reestimate the model T times if there are T periods in the sample) and then take the average of the coefficient estimates across all periods and test it against zero (Fama and MacBeth 1973). This approach avoids the look-ahead bias because, each time we estimate the coefficient, we do not use any information from the future in the covariance matrix.

[21] Past studies widely use return predictability to quantify the aggregated informativeness of a group of investors (Boehmer et al. 2008; Goetzmann and Kumar 2008; Hvidkjaer 2008; Kaniel et al. 2008; Grinblatt et al. 2012; Kelley and Tetlock 2013, 2016; Gao and Huang 2020).

[22] Online Table G4 presents robustness checks using retail order imbalance.

[23] 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). Online Appendix I presents the correlations of the variables in Table 7 and average variance inflation factors.

[24] Each participant is given a random number of either one, two, or three. Individuals with one are assigned to the treatment group, two to the control group, and three to the placebo group.

[25] Only 12 individuals (out of 100) in the treatment group did not click on the CSV flat file button.

[26] Among the 100 stocks, 95 were randomly selected, and the remaining five were selected from the top five gainers list. All stock data are real data from 2017.

[27] People with extremely low financial literacy are less likely to invest in stocks (Van Rooij et al. 2011).

[28] With \$7 for the 20-minute task, the hourly wage is \$21, which is higher than any state or federal minimum wage as of May 2022.

[29] In an untabulated table, we observe convergence between the demographic information in our sample and other studies. The percentage of females (45.39%) in our sample converges with that of the general retail investment population in 2017 (45.4%). Although subjects in our sample tend to be younger (partially reflecting the recent influx of young retail investors), their relative trading experience (ratio of investment experience to age) is comparable to that in prior studies (Dorn and Huberman 2005, Glaser and Weber 2007). Finally, the percentage of subjects with a college or university degree in our sample is 71%, similar to the 70% reported in Dorn and Huberman (2005).

[30] We conduct a back-of-the-envelope calculation to gauge the welfare implications of the API shutdown as follows. Our randomized experiment shows that the one-month return for subject participants with API-like data access is 3.8% lower (column (2) of Table 10). The quasi-natural experiment shows that the retail trading volume goes down by 8.6%–10.5%, on average, so we take the midpoint of 9.55%. Given that aggregate retail trades in our data were worth \$209 billion in 2017, retail investors as a whole saved 3.8% × 9.55% × 209 billion ≈ \$758 million for one year after the API shutdown, assuming every retail investor in the data set used the API. If we assume that 10% of retail investors in the data set were API users, then annualized saving would be \$75.8 million.

[31] See https://www.sec.gov/spotlight/fintech, https://www.sec.gov/investment/investor-alerts-and-bulletins/autolistingtoolshtm.

References

Amihud Y (2002) Illiquidity and stock returns: cross-section and time-series effects. *J. Financial Markets* 5(1):31–56.

Ammann M, Schaub N (2021) Do individual investors trade on investment-related internet postings? *Management Sci.* 67(9): 5679–5702.

Apesteguia J, Oechssler J, Weidenholzer S (2020) Copy trading. *Management Sci.* 66(12):5608–5622.

Barber BM, Odean T (2000) Trading is hazardous to your wealth: The common stock investment performance of individual investors. *J. Finance* 55(2):773–806.

Barber BM, Odean T (2001) Boys will be boys: Gender, overconfidence, and common stock investment. *Quart. J. Econom.* 116(1): 261–292.

Barber BM, Odean T (2002) Online investors: Do the slow die first? *Rev. Financial Stud.* 15(2):455–488.

Barber BM, Odean T (2007) All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Rev. Financial Stud.* 21(2):785–818.

Barber BM, Odean T (2013) The behavior of individual investors. Constantinides GM, Harris M, Stulz RM, eds. *Handbook of the Economics of Finance*, vol. 2, Part B (North Holland, Oxford, UK), 1533–1570.

Barber BM, Odean T, Zhu N (2009a) Do retail trades move markets? *Rev. Financial Stud.* 22(1):151–186.

Barber BM, Huang X, Odean T, Schwarz C (2022) Attention induced trading and returns: Evidence from Robinhood users. *J. Finance* 77(6):3141–3190.

Barber BM, Lee Y-T, Liu Y-J, Odean T (2009b) Just how much do individual investors lose by trading? *Rev. Financial Stud.* 22(2): 609–632.

Biais B, Hilton D, Mazurier K, Pouget S (2005) Judgmental overconfidence, self-monitoring, and trading performance in an experimental financial market. *Rev. Econom. Stud.* 72(2):287–312.

Blankespoor E, deHaan E, Zhu C (2018) Capital market effects of media synthesis and dissemination: Evidence from robo-journalism. *Rev. Accounting Stud.* 23(1):1–36.

Blankespoor E, deHaan E, Wertz J, Zhu C (2019) Why do individual investors disregard accounting information? The roles of information awareness and acquisition costs. *J. Accounting Res.* 57(1):53–84.

Boehmer E, Jones CM, Zhang X (2008) Which shorts are informed? *J. Finance* 63(2):491–527.

Boehmer E, Jones CM, Zhang X, Zhang X (2021) Tracking retail investor activity. *J. Finance* 76(5):2249–2305.

Bonsall SB IV, Green J, Muller KA III (2020) Market uncertainty and the importance of media coverage at earnings announcements. *J. Accounting Econom.* 69(1):101264.

Bursztyn L, Ederer F, Ferman B, Yuchtman N (2014) Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica* 82(4):1273–1301.

Bushee B, Cedergren M, Michels J (2020) Does the media help or hurt retail investors during the IPO quiet period? *J. Accounting Econom.* 69(1):101261.

Clarke J, Chen H, Du D, Hu YJ (2020) Fake news, investor attention, and market reaction. *Inform. Systems Res.* 32(1):35–52.

Conrad J, Kapadia N, Xing Y (2014) Death and jackpot: Why do individual investors hold overpriced stocks? *J. Financial Econom.* 113(3): 455–475.

Da Z, Engelberg J, Gao P (2011) In search of attention. *J. Finance* 66(5):1461–1499.

Da Z, Huang X, Jin LJ (2021) Extrapolative beliefs in the cross-section: What can we learn from the crowds? *J. Financial Econom.* 140(1): 175–196.

Dhar R, Zhu N (2006) Up close and personal: Investor sophistication and the disposition effect. *Management Sci.* 52(5):726–740.

Dorn D, Huberman G (2005) Talk and action: What individual investors say and what they do. *Rev. Finance* 9(4):437–481.

Dorn AJ, Dorn D, Sengmueller P (2015) Trading as gambling. *Management Sci.* 61(10):2376–2393.

Fama EF, MacBeth JD (1973) Risk, return, and equilibrium: Empirical tests. *J. Political Econom.* 81(3):607–636.

Farrell M, Green TC, Jame R, Markov S (2022) The democratization of investment research and the informativeness of retail investor trading. *J. Financial Econom.* 145(2):616–641.

Feng L, Seasholes MS (2005) Do investor sophistication and trading experience eliminate behavioral biases in financial markets? *Rev. Finance* 9(3):305–351.

Fernandes D, Lynch JG Jr, Netemeyer RG (2014) Financial literacy, financial education, and downstream financial behaviors. *Management Sci.* 60(8):1861–1883.

French KR (2008) Presidential address: The cost of active investing. *J. Finance* 63(4):1537–1573.

Frieder L, Subrahmanyam A (2005) Brand perceptions and the market for common stock. *J. Financial Quant. Anal.* 40(1):57–85.

Frydman C, Wang B (2020) The impact of salience on investor behavior: Evidence from a natural experiment. *J. Finance* 75(1):229–276.

Gao M, Huang J (2020) Informing the market: The effect of modern information technologies on information production. *Rev. Financial Stud.* 33(4):1367–1411.

Gao X, Lin T-C (2015) Do individual investors treat trading as a fun and exciting gambling activity? Evidence from repeated natural experiments. *Rev. Financial Stud.* 28(7):2128–2166.

Glaser M, Weber M (2007) Overconfidence and trading volume. *Geneva Risk Insurance Rev.* 32(1):1–36.

Goetzmann WN, Kumar A (2008) Equity portfolio diversification. *Rev. Finance* 12(3):433–463.

Greene J, Smart S (1999) Liquidity provision and noise trading: Evidence from the "investment dartboard" column. *J. Finance* 54(5): 1885–1899.

Grinblatt M, Keloharju M (2000) The investment behavior and performance of various investor types: A study of Finland's unique data set. *J. Financial Econom.* 55(1):43–67.

Grinblatt M, Keloharju M (2009) Sensation seeking, overconfidence, and trading activity. *J. Finance* 64(2):549–578.

Grinblatt M, Keloharju M, Linnainmaa JT (2012) IQ, trading behavior, and performance. *J. Financial Econom.* 104(2):339–362.

Grossman SJ, Miller MH (1988) Liquidity and market structure. *J. Finance* 43(3):617–633.

Hall CC, Ariss L, Todorov A (2007) The illusion of knowledge: When more information reduces accuracy and increases confidence. *Organ. Behav. Human Decision Processes* 103(2):277–290.

Hammond D, Fong GT, McNeill A, Borland R, Cummings KM (2006) Effectiveness of cigarette warning labels in informing smokers about the risks of smoking: Findings from the International Tobacco Control (ITC) Four Country Survey. *Tobacco Control* 15(suppl 3):iii19–iii25.

Han B, Tang Y, Yang L (2016) Public information and uninformed trading: Implications for market liquidity and price efficiency. *J. Econom. Theory* 163:604–643.

Huang S, O'Hara M, Zhong Z (2021) Innovation and informed trading: Evidence from industry ETFs. *Rev. Financial Stud.* 34(3):1280–1316.

Huffman D, Raymond C, Shvets J (2022) Persistent overconfidence and biased memory: Evidence from managers. *Amer. Econom. Rev.* 112(10):3141–3175.

Hvidkjaer S (2008) Small trades and the cross-section of stock returns. *Rev. Financial Stud.* 21(3):1123–1151.

Ivković Z, Sialm C, Weisbenner S (2008) Portfolio concentration and the performance of individual investors. *J. Financial Quant. Anal.* 43(3):613–655.

Kaniel R, Saar G, Titman S (2008) Individual investor trading and stock returns. *J. Finance* 63(1):273–310.

Kelley EK, Tetlock PC (2013) How wise are crowds? Insights from retail orders and stock returns. *J. Finance* 68(3):1229–1265.

Kelley EK, Tetlock PC (2016) Retail short selling and stock prices. *Rev. Financial Stud.* 30(3):801–834.

Kumar A (2009) Who gambles in the stock market? *J. Finance* 64(4):1889–1933.

Kumar A, Lee CM (2006) Retail investor sentiment and return comovements. *J. Finance* 61(5):2451–2486.

Langer EJ (1975) The illusion of control. *J. Personality Soc. Psych.* 32(2):311–328.

Lawrence A, Ryans J, Sun E (2017) Investor demand for sell-side research. *Accounting Rev.* 92(2):123–149.

Lee CM, Ready MJ (1991) Inferring trade direction from intraday data. *J. Finance* 46(2):733–746.

Lee CM, Zhong Q (2022) Shall we talk? The role of interactive investor platforms in corporate communication. *J. Accounting Econom.* 74(2–3):101524.

Liao L, Wang Z, Xiang J, Yan H, Yang J (2021) User interface and firsthand experience in retail investing. *Rev. Financial Stud.* 34(9):4486–4523.

Lou D (2014) Attracting investor attention through advertising. *Rev. Financial Stud.* 27(6):1797–1829.

Moore DA (2023) Overprecision is a property of thinking systems. *Psych. Rev.* 130(5):1339–1350.

Moore DA, Healy PJ (2008) The trouble with overconfidence. *Psych. Rev.* 115(2):502–517.

Newey WK, West KD (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3):703–708.

Ortoleva P, Snowberg E (2015) Overconfidence in political behavior. *Amer. Econom. Rev.* 105(2):504–535.

Oskamp S (1965) Overconfidence in case-study judgments. *J. Consulting Psych.* 29(3):261–265.

Ozik G, Sadka R, Shen S (2021) Flattening the illiquidity curve: Retail trading during the COVID-19 lockdown. *J. Financial Quant. Anal.* 56(7):2356–2388.

Peress J, Schmidt D (2020) Glued to the TV: Distracted noise traders and stock market liquidity. *J. Finance* 75(2):1083–1133.

Sias RW, Starks LT (1997) Institutions and individuals at the turn-of-the-year. *J. Finance* 52(4):1543–1562.

Statman M, Thorley S, Vorkink K (2006) Investor overconfidence and trading volume. *Rev. Financial Stud.* 19(4):1531–1565.

Stein K (2018) Remarks at SEC Speaks: Increasing product complexity: What's at stake? Accessed March 10, 2024, https://www.sec.gov/news/speech/stein-sec-speaks-increasing-product-complexity.

Svenson O (1981) Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica (Amsterdam)* 47(2):143–148.

U.S. Securities and Exchange Commission (2012) Study regarding financial literacy among investors. Report, U.S. Securities and Exchange Commission, Washington, DC.

Van Rooij M, Lusardi A, Alessie R (2011) Financial literacy and stock market participation. *J. Financial Econom.* 101(2):449–472.

Von Gaudecker H-M (2015) How does household portfolio diversification vary with financial literacy and financial advice? *J. Finance* 70(2):489–507.

Zhang XM, Zhang L (2015) How does the internet affect the financial market? An equilibrium model of internet-facilitated feedback trading. *Management Inform. Systems Quart.* 39(1):17–37.