This Time is Different: Investing in the Age of Robinhood

Valeria Fedyk *

November 2024

[Click here for the latest version]

Abstract

This paper studies the investing behavior and asset pricing implications of the new class of retail traders exemplified by the relatively young, small, and inexperienced individual investors on the Robinhood platform. I first document an absence of investing in lottery stocks, value stocks, and small-cap stocks, in contrast to prior literature, and show that these differences arise from a combination of investor behavior and market design changes such as the introduction of fractional share trading. I then identify three key drivers of Robinhood investment: a novel "buy-the-dip" effect, event-based trading in response to earnings announcements and analyst recommendation revisions, and trading connected with popularity and sentiment on the WallStreetBets platform. I develop a model to shed light on the buy-the-dip phenomenon and introduce a novel financial dictionary based on WallStreetBets sentiment. Finally, I analyze performance and show that Robinhood investment predicts returns and improves price discovery up to a one month horizon.

Keywords: Robinhood, WallStreetBets, retail trading, portfolio choice, investments, buy-the-dip effect, price discovery, FinTech

JEL codes: G11, G12, G14, G40, G41

*Arizona State University. **Email:** vfedyk@asu.edu. **Acknowledgments:** I owe my sincere gratitude for the invaluable mentorship of my advisors Anna Pavlova, Christopher Hennessy, and Suleyman Basak. I would also like to thank Svetlana Bryzgalova, Joao Cocco, James Dow, Alex Edmans, Julian Franks, Francisco Gomes, Roberto Gomez Cram, Marco Grotteria, Lakshmi Naaraayanan, Narayan Naik, Henri Servaes, and the other faculty at London Business School for thoughtful feedback. I am also grateful for the helpful guidance of Nicholas Barberis and the other faculty members at Yale University during my visit. I am thankful for insightful comments from Cecilia Bustamante, John Cochrane, Anthony Cookson, Anantha Divakaruni (discussant), Greg Eaton (discussant), Joseph Engelberg, Anastassia Fedyk, Matthew Gentzkow, Juanita Gonzalez-Uribe, Ahmed Guecioueur (discussant), Alexander Guembel, Nick Hirschey, Juan Imbet, Bryan Kelly, Hanno Lustig, Ken Singleton, Paolo Sodini, Laura Starks, David Stolin, Raman Uppal, Mindy Xiaolan, Paul Yoo (discussant), Irina Zviadadze, and participants at the Dauphine Finance PhD Workshop, Trans-Atlantic Doctoral Conference, IABF, WFC, LBS PhD Alumni Workshop, Helsinki Finance Summit on Investor Behavior, FMA, AFBC, SWFA, JARSiF, HEC Finance PhD Workshop, and Chicago Booth Asset Pricing Conference. The project also benefited from discussions with participants of the Robinhood and WallStreetBets platforms, who provided valuable insights into the institutional details.

1 Introduction

Recent years have witnessed a marked shift in the composition of U.S. equity traders. Over the past several decades, retail investors have tended to represent a relatively small proportion of total trading volume. In fact, as recently as 2019 only 15% of U.S. households held stocks directly and almost half were not invested in the stock market through any type of investment vehicle, precipitating the longstanding stock market participation puzzle.[1] However, the recent advent of zero commission, easy-to-use trading platforms such as Robinhood has contributed to a significant rise in the participation of retail investors in the U.S. stock market. While only 10% of trading volume came from retail traders in 2010, that number had doubled to 20% by 2021.[2] Furthermore, half of that remarkable increase—5% of the entire U.S. stock market volume—can be directly attributed to the rise of the generally young, small and inexperienced investors on the Robinhood platform. In this paper, I study the investing behavior of this new class of individual investors and document that their investments are markedly different from previously studied retail traders. I characterize their investments through the three main drivers of a novel buy-the-dip effect, event-based trading, and peer effects, and demonstrate that they positively predict future stock market returns up to a one month horizon.

The examination of this new class of retail investors is interesting and important for several reasons. First, while a substantial body of literature has explored the collective preferences of individual investors, there remains a knowledge gap regarding the behaviors of individual investors during the early stages of their investment journey.[3] The fact that the Robinhood platform has attracted a clientele characterized as young, small, and relatively inexperienced presents a unique opportunity for studying their investment behaviors.[4] Second, while the investing preferences of these investors are interesting in their own right, early-life experiences have also been shown to exert a lasting influence on financial attitudes and behaviors in the future (Malmendier, Tate, and Yan (2011), Bernile, Bhagwat,

[1]Source: Survey of Consumer Finances, 2019.

[2]Source: Bloomberg Intelligence. Furthermore, several financial institutions such as Credit Suisse place the percentage of retail trading volume as high as a third of the total volume at certain points of 2021. Retail trade volume is now almost as high as the volume from mutual funds and hedge funds combined.

[3]See Barber and Odean (2013) for an overview of the individual investor preferences literature.

[4]The median age of investors on the Robinhood platform is just 31 years old, the average account size is $3,500, and around half of the investors are first-time investors (source: Robinhood IPO S-1 Filing). These characteristics contrast sharply with the average previously studied retail investor. For instance, the average age of investors in the retail brokerage data used by Barber and Odean (2000, 2001, 2008) is 55 years old and the average account size is $35,629. Note that while the age, experience, and size characteristics Robinhood investors differentiate them from previously studied investors, Robinhood investors are in fact representative of the new class of retail investors as shown in Appendix H.

and Rau (2017)). This can become especially relevant once these investors command significantly more capital. Third, this new class of retail investors, to the surprise of many, has already had a significant impact on the financial markets.[5] Fourth, this study can help elucidate the impact of the rising retail class on price discovery and market efficiency, as well as inform policy discussions on safeguards for individual investor participation in the financial markets. Finally, the investing behaviors of the new young investor class have important consequences for household finance, which studies how consumption, savings, and investment decisions change over the life cycle.

To quantify Robinhood investment activity, I leverage the Robintrack dataset, which provides data on the number of Robinhood investors in each U.S. stock throughout the sample period spanning from May 2, 2018 to August 13, 2020.[6] My analysis is performed at a daily frequency and yields a comprehensive dataset with over 1,700,000 observations. Additionally, I explore data from the WallStreetBets platform, an investing subreddit frequented by Robinhood investors. To identify company names in WallStreetBets posts, I build and employ a company name matching algorithm, and to assess sentiment within these posts, I develop a novel sentiment dictionary for social media, including emojis.[7]

In the first part of the paper, I document the differences between Robinhood investors and previously studied individual investors. Motivated by the common media portrayal of Robinhood traders as gamblers, the first such major difference I identify is that Robinhood investors actually do not exhibit an outsized demand for investing in lottery stocks. In my analysis, I employ two established definitions of lottery stocks. The first definition, as proposed by Kumar (2009), categorizes lottery stocks based on three key characteristics: low price, high idiosyncratic volatility, and high idiosyncratic skewness. The second definition, based on Barberis and Huang (2008), designates lottery stocks as those displaying high idiosyncratic skewness. Employing both panel regressions and Fama and MacBeth (1973) regressions that include the lottery stock characteristics and control for conventional firm characteristics, I find that Robinhood investors do not invest highly in lottery stocks.

[5]As an example, consider the retail-trader-driven Gamestop short squeeze episode in January 2021, analyzed in more detail in Allen et al. (2023).

[6]I also investigate the growth of investment on the Robinhood platform after 2020 and find that the platform—both as a measure of net cumulative funded accounts and net deposits by investors—continued to grow in the period 2020–2023. This analysis is provided in Appendix G and demonstrates that the new class of retail investors does not appear to be a temporary phenomenon and is worth a more thorough investigation.

[7]Appendix B outlines the key data processing steps. The company name matching algorithm and sentiment dictionary are discussed in Section 4.3.

I explain the absence of lottery stock investment through a well-identified market design change. Specifically, in December 2019 the Robinhood platform introduced fractional share trading. This represented an impactful intervention for many of the previously capital-constrained investors on the platform, since the average account size in Robinhood is only $3,500. As a result, Robinhood investors who faced binding constraints in buying high nominal price stocks prior to the introduction of fractional share trading benefitted from the opportunity to invest in fractional shares after the implementation of this change.[8] In order to test whether the introduction of fractional share trading impacted Robinhood investment in lottery stocks, I employ a regression discontinuity in time (RDiT) design around the introduction date of December 12, 2019. I find that the introduction of fractional share trading is associated with a 32 percentage point increase in the average loading on stock price among Robinhood investors. These results solidify the connection between the adoption of fractional share trading and the decline in investment in lottery stocks by the emerging retail investor class. They also demonstrate how platform design can help mitigate investor biases. In this case, for example, the market design solution relaxed the financial constraints of individual investors and ameliorated their previously suboptimal investment decisions.

How else are Robinhood investors different from what we know about retail investors? Two additional differences I identify pertain to Robinhood investors' investment patterns in value stocks and small-cap stocks. I find that contrary to a propensity for value and small-cap investments documented in prior literature (Barber and Odean (2000)), Robinhood investors actually exhibit a significantly higher inclination towards large-cap and growth stocks. Specifically, I document that a one standard deviation increase in size (growth) corresponds to a 53bps (26bps) increase in Robinhood investment the following day. In order to explain these findings, I analyze value and size factor performance in the years leading up to my Robinhood sample period and in the years preceding the individual investor sample analyzed by Barber and Odean (2000). My analysis reveals evidence that is consistent with extrapolative beliefs of retail investors in the domain of style investing, which is a novel finding.

After documenting that Robinhood investors behave differently from several previously documented patterns, in the second part of the paper I turn to identifying what does drive the investing behavior of this new retail class. I am able to identify three key drivers: a novel buy-the-dip effect, event-based

[8]In fact, Robinhood reports that in the first quarter of 2021 alone, over 40% of customer equity trades by number were fractional trades (source: Robinhood IPO S-1 Filing).

trading, and peer effects through WallStreetBets.

The first of these three key drivers is a novel phenomenon which I term the buy-the-dip effect. For large, well-known companies, Robinhood investors have a predilection for buying after an extreme negative return, or "buying the dip." While the buy-the-dip effect is present for large-cap companies, it largely disappears for small-cap companies. This pattern is also robust to alternative measures of negative events, such as bottom quintile earnings surprises. I propose a hypothesis to explain this behavior, positing that investors wait for the "right moment" to buy well-known companies, which tend to be large-cap, and perceive the right moment as occurring after the company's price has experienced a dip.

To formalize this intuition, I develop a belief-based model that encapsulates this logic and is able to explain the buy-the-dip effect. In the model, individual investors suffer from the well-known gambler's fallacy for large-cap stocks.[9] When returns for a large-cap stock are negative, investors attribute a larger portion of the decline to noise rather than a true signal reflecting the firm's underlying fundamentals. Consequently, they anticipate future mean reversion in the stock price and increase their investment in the stock. Conversely, for small-cap stocks investors update on realized returns by treating them more as signal than noise. The implications are that individual investors tend to engage in buy-the-dip behavior for large-cap companies, while they do not exhibit this investment pattern for small-cap stocks.

The second of the three main components of Robinhood investor behavior that I identify is event-based trading, indicating that the new class of retail investors is sensitive to corporate announcements and financial analyst reports. In particular, I find that Robinhood investment is five times higher for stocks on earnings announcement days relative to control no-announcement days, and six times higher for stocks on analyst recommendation revision days versus control no-revision days. Upon closer examination of announcement days, I find that conditional on an announcement being made Robinhood investment increases more if the information released is extreme. For example, days with earnings surprises in the top and bottom quintiles exhibit significantly higher Robinhood investment than days in the earnings management region. Additionally, I also note that Robinhood investment increases in response to high volume traded, high volatility, and extreme returns. I explain these findings using the

[9]The gambler's fallacy and hot hands fallacy can both be traced back to the law of small numbers behavioral bias, which is described in Kahneman and Tversky (1971).

concepts of bounded rationality and limited attention (Grinblatt and Keloharju (2000), Seasholes and Wu (2007), Barber and Odean (2008)). I rule out an alternative explanation centered on the private information channel, as I find no abnormal pre-announcement drift in Robinhood investment in the days leading up to earnings announcements or analyst recommendation revisions.

To identify the third and final driving force behind Robinhood investment behavior, I conduct an analysis of posts on the investment advice subreddit, WallStreetBets. This subreddit was started within 15 months of Robinhood's inception, and has garnered significant popularity among Robinhood users. My findings reveal a substantial increase in Robinhood investment for stocks that exhibit higher levels of popularity and more bullish sentiment on the WallStreetBets platform. Specifically, the number of mentions of a stock, the number of comments on posts discussing the stock, the aggregate net upvotes of those posts, the aggregate awards given to those posts, and the sentiment of those posts all significantly predict next-day increases in Robinhood investment. This underscores that, in addition to the buy-the-dip effect and event-based trading, Robinhood investment is also influenced by the collective sharing of information, opinions, and sentiment within the WallStreetBets platform. Moreover, I underscore a novel information diffusion channel based on slang terminology and emojis. Specifically, in order to measure the sentiment of both written vocabulary and graphical representations, I implement a simple, straightforward approach to separate bullish and bearish posts, and use this to construct my own WallStreetBets sentiment dictionary.[10] This allows me to systematically compute a daily sentiment measure for each stock discussed on the platform, which can also be used in future research.[11]

In the final segment of my paper, I assess the financial market implications stemming from the emergence of this new class of retail investors. I construct two representative portfolios of Robinhood investors, one employing the dollar-weighted method and the other relying on the share-weighted method.[12] Under both methodologies, I observe a substantial annualized alpha of approximately 17%. This outperformance remains statistically significant under the Fama-French three-factor, Fama-French-Carhart four-factor, and Fama-French six-factor models. Breaking out the performance by year, I find that the majority of this outperformance materialized in 2020, although Robinhood investors also

[10] To my knowledge, this is the first such financial dictionary to analyze WallStreetBets slang phrases and graphical emojis. I also find that this approach achieves higher post categorization accuracy than methods based on bag-of-words, word embedding, and support vector machine (SVM) machine learning algorithms.

[11] The WallStreetBets sentiment dictionary is available for academic research use at https://www.valeriafedyk.com/data.

[12] The construction of the representative portfolios follows Welch (2022).

outperformed in 2019. I then decompose performance according to the three key components underlying Robinhood investment that I identified earlier. My analysis reveals that both the buy-the-dip effect and event-based trading strategies contributed significantly to the outperformance observed over the sample period, while a portfolio based on WallStreetBets did not yield statistically significant returns. These findings underscore the possibility that the new class of retail investors may be benefiting from the risk premium associated with announcement event days (Savor and Wilson (2014), Savor and Wilson (2016)), even as their investment behaviors may be influenced by behavioral effects such as bounded rationality, limited attention, and the law of small numbers. Lastly, I explore the return predictability associated with Robinhood trades and find compelling evidence that Robinhood investment predicts returns up to a one month horizon.

The results in this paper offer novel insights into the investing behavior of the emerging class of retail traders. The absence of investment in lottery stocks, value stocks, and small-cap stocks underscores the distinctions that set these investors apart. Additionally, these findings shed light on how platform design can help mitigate individual investor biases.

These results are also relevant for the decision-making of institutional investors and supply-side market participants, including corporate management. The responses of the burgeoning new class of retail traders to firm-specific announcements, such as earnings releases and analyst recommendation revisions, serve as indicators that this group is highly responsive to such corporate disclosures. Moreover, the outperformance observed not only in the event-based trading strategy but also of the novel buy-the-dip effect may pique the interest of institutional investors and companies facing challenging circumstances, as they seek to attract new capital infusions through equity investment.

Finally, the results also speak to the financial market effects of the increase in the new class of retail traders in the U.S. over the past five years. Insofar as the return predictability exhibited by Robinhood investors remains robust in future periods, the presence of such investors has the potential to accelerate the pace of price discovery and enhance overall market efficiency. Moreover, the favorable performance of these investors during the sample period, coupled with their sustained growth, carries notable policy implications for encouraging stock market participation and reassessing individual investor safeguards in the years ahead.

Related Literature

My paper merges insights from financial economics, behavioral finance, and asset pricing and advances three main strands of literature. The first is on the investing preferences and behavioral biases exhibited by individual investors; the second is the emerging line of inquiry into Robinhood traders and the new class of retail traders as a whole; and the third is the literature on sentiment analysis, information diffusion, and price discovery. To the best of my knowledge, this paper is the first to introduce the buy-the-dip effect and develop a unified framework to explain it. This paper also differs from previous work by uncovering an absence of lottery stock investing behavior among the new retail class, and explaining it via a well-identified platform design change. Finally, the paper contributes to the sentiment and information diffusion literatures by proposing a novel sentiment dictionary based on new-age slang terminology and graphical representations and to the price discovery literature by documenting positive return predictability of Robinhood investment up to a one month horizon.

Seminal studies on the investing preferences of individual investors include Odean (1999) and Barber and Odean (2000, 2001), who find that individual investors are overconfident, engage in excessive trading, and ultimately underperform. They also document that these investors tend to overweight high beta, small-cap, and value stocks. Kaniel, Saar, and Titman (2008), Kelley and Tetlock (2013), and Barrot, Kaniel, and Sraer (2016) find that retail trades predict returns in a way consistent with liquidity provision, while Foucault, Sraer, and Thesmar (2011) document that individual investors increase volatility and behave as noise traders. Kaniel et al. (2012) and Luo et al. (2022) investigate contrarian trading behavior and the impact of retail trading on post-earnings-announcement drift. Kumar (2009) demonstrates that retail investors have a propensity for investing in lottery stocks, and also documents the underperformance of these investors. Garrett and Sobel (1999) explain the participation of risk-averse individuals in state lotteries through a preference for skewness, while Barberis and Huang (2008) find a predilection for lottery-type investment among retail traders using idiosyncratic skewness to define lottery stocks. Bali, Cakici, and Whitelaw (2011) and Amaya et al. (2015) provide further analyses of lottery stocks. What sets my paper apart is the revelation that Robinhood investors deviate from the established norms on several dimensions: they do not have high investment in lottery stocks, value stocks, or small-cap stocks. In addition, I document a novel buy-the-dip effect that helps explain both the investment choices of the new retail class and their patterns of trading around earnings announcements. Furthermore, a representative portfolio of Robinhood investors—driven by event-based trading and the buy-the-dip effect—significantly outperforms the broader market, further distinguishing their investment behavior.

As Robinhood investors have taken the popular media by storm, they have also garnered more attention from academics. Studies of Robinhood investors include investigations into ESG preferences (Moss, Naughton, and Wang (2023)), liquidity effects during the Covid-19 pandemic (Ozik, Sadka, and Shen (2021)), intraday reactions to price changes (Ardia, Aymard, and Cenesizoglu (2023)), information salience (Stein (2020)), retail demand curve estimation (van der Beck and Jaunin (2023)), and market quality (Pagano, Sedunov, and Velthuis (2021), Eaton et al. (2022)). Of particular relevance to my research, Welch (2022) documents the outperformance of Robinhood investors and their affinity for high-volume stocks, while Barber et al. (2022) reveal that herding episodes on the Robinhood platform can be predicted by attention measures (such as recent investor interest, extreme returns, or unusual volume) and predict negative abnormal returns. In my paper I concentrate on the drivers behind Robinhood outperformance and identify three key novel features that help explain Robinhood investment: the buy-the-dip effect, event-based trading, and peer effects through WallStreetBets. To my knowledge, my paper is the first to empirically identify and provide a theoretical mechanism for the buy-the-dip effect for large-cap stocks. Additionally, this paper is the first to link analyst recommendation revisions with Robinhood trading behavior and event-based trading around earnings announcements with their outperformance. Furthermore, I also build a unique sentiment dictionary based on WallStreetBets, which adds to the body of literature focused on sentiment and textual analysis (e.g., Loughran and McDonald (2011), Cookson et al. (2022)). This dictionary is simple and easy to use, and provides a way of deciphering the new-age lexicon that relies heavily on emojis and slang terminology.

Related literature also includes the burgeoning study of the new class of investors as a whole, across a wide variety of trading platforms and security types. Recent work within this domain has focused on the effects of stimulus checks on retail investing (Greenwood, Laarits, and Wurgler (2023)), the identification of retail trades using TAQ data (Boehmer et al. (2021)), the impact of fractional share trading on meme stock-like trading frenzies (Da, Fang, and Lin (2023)), retail investing in hard-to-value stocks (Laarits and Sammon (2023)), the heterogeneity of investor types on WallStreetBets (Uettwiller (2022)), social network coordination efforts surrounding meme stocks (Pedersen (2022), Allen et al. (2023), Hu et al. (2023)), smartphone investing (Kalda et al. (2021)), and investment in options (Beckmeyer, Branger, and Gayda (2023)) and cryptocurrencies (Kogan et al. (2023)).

My research also contributes to the behavioral finance literature. Over the past few decades, numerous studies have shed light on the phenomena of limited attention among investors and their tendency to gravitate toward attention-grabbing securities (Grinblatt and Keloharju (2000), Seasholes and Wu (2007), Barber and Odean (2008)). Other studies have delved into the realm of cognitive biases, such as the small sample bias (Kahneman and Tversky (1971)), and extended their investigations to encompass concepts such as the gambler's fallacy (Rabin (2002)) and the hot hands fallacy (Rabin and Vayanos (2010)). My contribution to this area lies in linking the three key components of Robinhood investment that I identify with these behavioral finance mechanisms. In particular, I provide a model for the buy-the-dip effect based on the gambler's fallacy and link event-based trading and peer effects on WallStreetBets with theories of limited attention.

Finally, my paper also adds to the literature on sentiment analysis, price discovery and information diffusion (e.g. Tetlock (2007), Da, Engelberg, and Gao (2011), Fedyk (2022), Cookson, Engelberg, and Mullins (2023), Garcia, Hu, and Rohrer (2023)). I provide empirical evidence of return predictability and enhanced price discovery arising from Robinhood investment. Moreover, using a novel dictionary that differs significantly from traditional media, I establish a connection between the sentiment of topics discussed on the WallStreetBets platform and subsequent investment choices made by Robinhood investors. I demonstrate that popularity has a stronger effect than sentiment, but both aspects of discussions are significant predictors of Robinhood investment decisions.

The remainder of the paper is organized as follows. In Section 2, I describe the data used in this paper. Section 3 presents my main findings regarding the differences in investing preferences of the new class of Robinhood investors. Section 4 explores the buy-the-dip effect, event-based trading, and how the WallStreetBets social network influences Robinhood traders' investment decisions. Section 5 measures Robinhood representative portfolio performance against several asset pricing model benchmarks, decomposes it over time, and relates it to the three key drivers. Section 6 discusses policy implications. Finally, Section 7 concludes.

2 Data

In this section, I describe the data sources and methodology used to construct the full sample dataset. I provide descriptive statistics for the coverage of Robinhood securities in CRSP, the breakdown of Robinhood securities by type, the collective diversification of Robinhood holdings, the top holdings among Robinhood investments, and the sector weights of the Robinhood portfolio versus the market portfolio according to the SIC and NAICS industry classification systems. Finally, I investigate the impact of the Covid-19 pandemic shock and show that it coincided with rapid Robinhood investment growth and that Robinhood has continued to grow through 2023.

2.1 Robintrack

The primary data source used to analyze retail investors in this paper is the Robintrack dataset. Launched in 2018 by Casey Primozic, Robintrack collected hourly data on the number of investors holding each security on the Robinhood platform from May 2018 to August 2020. Amid concerns that this data might mischaracterize the company as pandering to day traders or otherwise expose it to negative media scrutiny, Robinhood ceased providing investor count data in August 2020. As a result, the dataset's coverage ends on August 13, 2020, which marks the final full day of available data.

During the sample period from May 2, 2018, to August 13, 2020, the Robintrack dataset provides the number of Robinhood investors in a given security at an hourly frequency. For this study, I aggregate the data to a daily frequency. Specifically, I define the number of investors in a stock S on day t as the last recorded number of investors in S prior to the close of trading on that day (4:00 PM EST). After filtering out intraday observations, duplicate share class tickers, and securities with no Robinhood investors throughout the sample period, the final dataset comprises 8,507 securities and date-security observations.[13]

An important limitation of the dataset is that it reports the number of investors rather than the dollar amount invested and does not identify which investors are trading a given security. Instead, it provides the aggregate number of Robinhood investors holding each security over time. Despite this, the dataset offers valuable insights into stock popularity and allows for inferences about investing behavior, assuming that the number of Robinhood investors in a security is representative of the dollar amount invested by Robinhood investors in that security. A number of studies on Robinhood performance, investment behavior, ESG press release reactions, and market quality have successfully leveraged this data to make significant contributions to the literature.[14]

After constructing the Robinhood dataset, I map the Robinhood tickers to CRSP permnos. The

[13] More details on the data cleaning procedures and sample construction can be found in Appendix B.1.

[14] See, for instance, Ardia, Aymard, and Cenesizoglu (2023), Barber et al. (2022), Da, Fang, and Lin (2023), Eaton et al. (2022), Moss, Naughton, and Wang (2023), Stein (2020), and Welch (2022).

details for this mapping are discussed at length in Appendix B.2. The analyses in this paper use these distinct permnos as the primary security identifiers.

I next assess the extent to which Robinhood securities can be mapped to CRSP and vice versa. Both mappings result in a high level of coverage, as evidenced by the coverage figures discussed below.[15] This extensive coverage is significant for two reasons. First, the broad representation of CRSP securities on Robinhood indicates that investors are not constrained by limited investment options on the platform, addressing a potential concern that Robinhood investments are driven by availability rather than choice. Second, the extensive coverage of Robinhood securities in CRSP data is important since it allows for a fuller analysis of investment in those securities. Through CRSP, I am able to match the vast majority of securities that Robinhood investors hold to their corresponding stock returns and other characteristics, which enables me to better analyze those holdings.

Figure 1 illustrates the coverage of Robinhood securities in CRSP. The blue line represents the total number of securities in Robinhood over the sample period, the orange line represents the number of Robinhood securities successfully mapped to CRSP, and the green line represents the number of Robinhood securities mapped specifically to CRSP U.S. stock securities. Of the 8,507 cleaned Robinhood tickers, 8,040 are successfully mapped to 7,969 unique permnos. Among these, 3,775 Robinhood tickers correspond to 3,722 unique U.S. stock permnos. The figure demonstrates that nearly all Robinhood securities can be matched to CRSP, with close to 3,000 securities mapped to U.S. stocks on any given date during the sample period. Overall, more than 95% of Robinhood tickers are successfully linked to CRSP.

The coverage of CRSP stocks in Robinhood is shown in Figure 2. In this direction as well, coverage is substantial. Throughout the entire sample period, over 80% of CRSP U.S. stocks can be found in the Robinhood sample, and coverage increases to over 95% by the end of the sample period. When I compute the aggregate market cap of all the CRSP U.S. stocks that can be mapped to Robinhood as a percentage of the total aggregate market cap of all U.S. stocks in CRSP, the coverage increases to over 90% over the course of the sample period and to over 97% by the end of the sample period. Consequently, investors are able to invest in nearly any U.S. stock security through the Robinhood platform.

[15] For the purposes of all illustrative coverage figures, I forward fill the number of Robinhood investors. Please see Appendix C.1 for more details on the forward filling procedure.

Figure 3 depicts the Robinhood coverage of U.S. stocks in CRSP broken down by market cap, where small cap includes all observations for stocks with a market cap less than or equal to $2bn, mid cap includes all observations for stocks with a market cap between $2bn and $10bn, and large cap includes all observations for stocks with a market cap equal to or over $10bn. The figure demonstrates that large-cap stocks have the best coverage, although coverage across all market cap sizes is very good. This indicates that studying the Robinhood dataset allows us to more accurately zero in on the investing preferences of the Robinhood investors, since they are hardly limited by the platform in terms of the securities they can invest in.

In terms of the security types traded on Robinhood, as shown in Panel (a) of Figure 4, almost half of the mapped securities available for trading in Robinhood are U.S. stocks with a further 29% of ETFs. A minority of securities fall under other security types: approximately 8% are foreign stocks, 7% are U.S. mutual funds, 5% are ADRs, 3% are REITs and 1% are various other types of securities. This demonstrates both the significant breadth of securities offered on the Robinhood platform and the fact that most of the securities on Robinhood are either U.S. stocks or ETFs. Notably, while Robintrack data includes a breadth of security types, it does not include data on the equity options or cryptocurrencies traded by Robinhood investors.[16]

Figure 4 Panel (b) illustrates the security type breakdown scaled by the aggregate number of Robinhood investors invested in each security type. To do this I rely on data from August 13, 2020, the final date of the sample. Specifically, I first sum the number of Robinhood investors in each security type and then display this breakdown in a pie chart as a percentage of the total sum of Robinhood investors across all security types on August 13, 2020. This breakdown reveals that a majority of the number of investors on the Robinhood platform, over 70%, are coming from investments in U.S. common stock. In this paper, I focus my analysis on this majority, namely the Robinhood investments in U.S. stocks.

2.2 WallStreetBets

The second relatively non-standard dataset I use is WallStreetBets. Created in 2012, WallStreetBets is a popular investing subreddit that has gained popularity among Robinhood investors, and is credited

[16] Robinhood introduced stock and ETF options trading on its platform in December 2017, and cryptocurrency trading over 2018-2021 in every state except Hawaii and Nevada.

by many as having been one of the organizing platforms for the Gamestop rally of January 2021. With over 14 million subscribers as of May 2024, the investing platform hosts numerous active stock market discussions each day. Anecdotal survey evidence of Robinhood participants indicates the platform is a source of trading information and trading signals for the Robinhood investors; I test and confirm this hypothesis in Section 4.3.

The dataset I use in my analysis covers all posts on the WallStreetBets subreddit over the period May 2, 2018 - April 30, 2020. The data comes from pushshift.io, and is not available following April 2020; nevertheless, most of the May 2018 - August 2020 Robinhood data sample falls under the period with WallStreetBets coverage. Each post includes the post title, post content, date posted, number of comments, number of awards, and the post score. Awards are online tokens of appreciation given by individual users to posts they particularly value, while the post score is the number of net upvotes that a post receives. Appendix B.4 provides more detail on the WallStreetBets dataset.

While it offers a valuable source of information on how Robinhood investors connect with each other and share investing tips, the WallStreetBets data is not without its own limitations. The primary such limitation is that most of the posts only contain the title but not the content of the post. The reason for this is that the post content is no longer available on the WallStreetBets forum, and is thus marked both in the data and on the website as removed. I hypothesize that this could frequently be due to outside images and gifs that no longer link to a valid url. Therefore, the analysis in such cases can only rely on the tickers and sentiment data available from the post title itself.

The second limitation is that posts frequently contain an image or a video. The name of the image or video is available in the textual data, but the image or video itself is not. The images and videos are not parsed or visually analyzed, and this direction of visual media sentiment analysis is left for future research.

Despite these minor drawbacks, the WallStreetBets data offers a wealth of information on retail investor stock market conversations, preferences, and ways of thinking about investing.

2.3 Other Data Sources

Alongside Robintrack and WallStreetBets, I also rely on several additional data sources. For each stock in the sample, I obtain the daily ticker, company name, permno, share code, exchange, Standard Industrial Classification (SIC) code, North American Industry Classification System (NAICS) code, price, bid price, ask price, return, share volume traded, and shares outstanding data from the Center for Research on Security Prices (CRSP). Stock fundamentals and accounting data used to compute the gross profitability, book-to-market ratio, and investment variables come from Compustat. I use analysts' quarterly earnings estimates and realized earnings from Thomson Financial's Institutional Brokers Estimate System (I/B/E/S) summary files to compute earnings surprises. Analyst recommendations are sourced from the I/B/E/S Recommendations - Detail file, and analyst recommendation revisions for each date and security are computed as the net number of upgrades less the number of downgrades by all analysts for the security on the given date. Factor returns for the value and size factors of Fama and French (1992) and the momentum factor of Carhart (1997) are sourced from Kenneth R. French's data library.[17] More details on the variables used and their construction can be found in Appendix A.

2.4 Descriptive Statistics

A popular belief surrounding Robinhood investors is that they are homogeneous and collectively undiversified. For instance, a typical stereotype might include visualizing all Robinhood traders as invested in a few particularly popular stocks such as Tesla. To investigate the validity of this assumption, I construct a histogram depicting the distribution of Robinhood investors across different stocks.

Figure 5 depicts this histogram. The histogram has a significant right skew (a very small number of stocks with a very large ownership base), so for illustrative purposes the figure is cut at the 90th percentile. Overall, the number of stocks with a very large number of Robinhood investors is very small. For instance, on the right tail of the distribution there are only eight stocks with over 500,000 Robinhood investors and 379—around 10% of the total stock universe—that have more than 10,000 Robinhood investors. On the other hand, there is a large number of stocks with a very small number of investors. A total of 1,393 stocks—which constitutes nearly 40% of all stocks in the universe—each have fewer than 500 Robinhood investors. Consequently, not all Robinhood investors are invested in the same stocks and there is considerable diversity among the Robinhood investors' collective holdings.

One natural question to arise from the Robinhood investors per stock histogram is what high-level probability distribution for stock selection would fit the observed data well. In a simulation exercise, I am able to generate a histogram with similar properties in the following manner. The parameters of

[17] The Kenneth R. French data library is available at: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

the simulation include 5,000 investors, a stock universe containing approximately 3,500 stocks (which is the number of U.S. stocks with a positive number of Robinhood investors and non-missing market cap values as of the sample period end date 8/13/2020), and a portfolio of ten stocks per investor. The probability that an investor chooses a given stock as one of the ten stocks in their portfolio is equal to the stock's market weight, i.e. the market cap of the stock divided by the market cap of the entire stock universe.

Figure 6 plots the histogram for the number of investors per stock following the rules of this simulation. The resulting figure possesses a number of similar properties to the original, notably a large number of stocks with very few Robinhood investors, a significant right skew, and a small number of stocks with a very large number of Robinhood investors. On the other hand, an alternative simulation using an equal weight probability distribution for stock selection does not match the data well. This indicates that the probability that a stock enters into a Robinhood investor's portfolio is likely positively related to its market cap. In turn, this suggests that in contrast to previously studied retail investors, Robinhood investors do not invest disproportionately in small-cap stocks (which I more formally demonstrate to be the case in Section 3.3). The selection process may be consistent with attention-induced trading, since large-cap companies tend to elicit more news and media mentions that garner attention. I show that Robinhood investment indeed responds positively to the popularity of stock discussions on the WallStreetBets investing subreddit in Section 4.3.1.

Figure 7 shows analogous histograms of the number of Robinhood investors per stock for small-cap, mid-cap, and large-cap stocks separately. Small cap is defined as any stock with a market cap of less than or equal to $2bn, mid cap is defined as any stock with a market cap between $2bn and $10bn, and large cap is defined as any stock with a market cap greater than or equal to $10bn as of the sample period end date. The pattern of the number of investors per stock is similar to the overall histogram, with a significant right skew and a diversity of holdings among Robinhood investors for each of the size category stocks. This demonstrates that the breadth of stocks collectively held by Robinhood investors is not limited to large-cap stocks, but rather extends to mid-cap and small-cap stocks as well.

Note that the preceding analyses provide evidence of large cross-sectional dispersion in the stocks held by Robinhood investors, and indicate that Robinhood investors as a group are invested in a diverse set of stocks. However, it does not appear to be the case that the average individual Robinhood investor is very well diversified. In fact, a quick back-of-the-envelope calculation reveals the opposite. As of the last day of the sample period, August 13, 2020, the total number of Robinhood investors across all U.S. stocks was 29,258,764 . At the same time, there was a total of approximately 9.725 million funded accounts on Robinhood.[18] Consequently, each account holder held an average of 29,258,764 / 9.725 million $\approx$ 3.01 stocks. This indicates that while the collective holdings across all Robinhood holders are very diverse, the average individual Robinhood investor does not appear to be well diversified.[19]

I also look at the most popular stocks among Robinhood investors. Table 1 depicts the top 10 holdings of Robinhood investors as of August 13, 2020, the sample end date. Note that the top 10 holdings are limited to U.S. stocks, and therefore do not include popular experience stock holdings from foreign companies or other security types, such as the Cambria Cannabis ETF. The top 10 holdings cover a number of large technology companies including Apple, Microsoft, and Amazon as well as six companies that illustrate the buy-the-dip effect described in Section 4.1. These six companies are the automakers Ford and Tesla, industrial conglomerate General Electric, technology company GoPro as well as the two airlines American Airlines and Delta Air Lines. Despite belonging to different sectors, what these companies all have in common is that each experienced a drawdown of over 50% during the Robinhood sample period of May 2, 2018 - August 13, 2020. These stocks are highlighted in blue in the table and discussed further in Section 4.1.

Finally, the sector weights of the Robinhood portfolio versus the market portfolio are reported in Table 2. Panel A shows the sectors denoted using the Standard Industry Classification (SIC) system, and sorted into ten top-level divisions following the classification of the U.S. Department of Labor's Occupational Safety and Health Administration.[20] U.S. stocks that fall into the "Nonclassifiable Establishments" group are given a classification code of -99. For the purposes of Table 2, such non-classified stocks are excluded so that the remaining percentage weights sum to one. One finding that emerges from considering the weights of the market and Robinhood portfolios in the unclassified stocks category is that the unclassified weight for the Robinhood portfolio is significantly higher than that of the market portfolio. Upon closer inspection, I find that the top five stocks in Robinhood that fall

[18]I arrive at this figure via interpolation between the 2019 and 2020 year-end numbers of net funded accounts reported by Robinhood. In particular, there were 5.1 million funded accounts as of December 31, 2019 and 12.5 million funded accounts as of December 31, 2020, for a total of approximately 5.1 + 7.5/12(12.5 - 5.1) = 9.725 million funded accounts as of mid-August 2020. Source: https://investingintheweb.com/brokers/robinhood-statistics/.

[19]When ETFs, foreign stocks, ADRs, REITs, mutual funds, and other security types are included in the analysis, the average number of securities held by a Robinhood investor rises to 4.30 . This is based on the total number of Robinhood investors across all security types of 41,841,388 and 9.725 million funded accounts on Robinhood.

[20]The full classification of SIC division codes is available at https://www.osha.gov/data/sic-manual.

into the unclassified establishments category according to SIC are Tesla (TSLA), GoPro (GPRO), Moderna (MRNA), Zynga (ZNGA), and Nikola (NKLA). These stocks include large, high growth, and technological companies, which gives credence to a recently popularized concern that the SIC system, having been developed for traditional industries prior to 1970, does not adequately recognize the newly emerged computer, software, and information technology sectors.[21] To circumvent this issue, Panel B reports the sector weights of the Robinhood portfolio and market portfolio using the North American Industry Classification System (NAICS). Stocks with unknown classification are not included in the sector weight computation.[22] For completeness, industry weights using the Fama French 49 Industry Classification are reported in Table IA.8.

The sector weights reported in Table 2 are computed as of the sample period end date of August 13, 2020, and show that Robinhood investment tended to be lower than the market in sectors such as finance and insurance, agriculture, and construction. Meanwhile, according to the NAICS classification, Robinhood investment was higher than the market in manufacturing and real estate. Note that I use two different methods for constructing the representative Robinhood portfolio: the dollar method and the share method. The rationale for using these two methods is that the Robintrack data only provides the number of Robinhood investors in each stock, rather than the actual dollar amount invested. Consequently, to construct a representative Robinhood portfolio it is necessary to make some assumptions regarding how much investment a Robinhood investor in a stock represents.[23]

The dollar and share method can be summarized as follows. In the dollar method, every Robinhood investor in a given stock is treated as an equal dollar amount investment in that stock. Without loss of generality, I let each investor represent a $1 investment in that stock. For instance, if stock X has 1,000 Robinhood investors the dollar method would treat stock X as having a $\$1,000$ aggregate investment from Robinhood investors. Once such dollar investments in each stock are calculated, the portfolio weight of a stock in the representative Robinhood portfolio can be computed as the Robinhood dollar investment in that stock divided by the total Robinhood dollar investment in all stocks. In this way, the portfolio weights sum to one. In the share method, each Robinhood investor in a stock is treated as an equal number of share investments in the stock. Without loss of generality, I let each Robinhood

[21] Source: https://en.wikipedia.org/wiki/Standard_Industrial_Classification.

[22] Such unclassified stocks comprise only 3.2% of the market portfolio. The corresponding figures are 7.4% for the dollar method Robinhood portfolio and 1.6% for the share method Robinhood portfolio.

[23] Welch (2022) similarly employs the dollar method and share method of building a representative Robinhood portfolio, though he does not apply it to the question of determining representative portfolio sector weights.

investor in a stock represent a 1 share investment in that stock. For instance, if stock Y has 1,000 Robinhood investors, the share method would dictate that the aggregate Robinhood portfolio holds 1,000 shares of stock Y. Once the number of shares for all stocks in the aggregate Robinhood portfolio are thus computed, I calculate the market value invested in each stock by multiplying the number of shares by the stock price. Finally, to arrive at the Robinhood representative portfolio weights I divide the Robinhood market value invested in a given stock by the aggregate market value of Robinhood investments in all stocks.

Although in this paper I focus on investment in U.S. stocks, for completeness I also provide descriptive statistics for Robinhood holdings in other types of securities in Appendix E.1.

2.5 The Covid-19 Pandemic Shock to Retail Ownership

Finally, I also document that the Covid-19 pandemic led to a significant increase in the number of investors on the Robinhood platform. Figure 8 reflects this increase in the following manner. The blue line represents the number of Robinhood investors in the S&P 500 ETF over time, while the orange line shows the average number of Robinhood investors in a top 30 Dow Jones stock over time.[24] The black dashed line represents February 3, 2020, which is the day the U.S. declared Covid-19 to be a public health emergency. Declarations of a state of emergency in each of the 50 states followed soon after, with Washington leading the way with a February 29, 2020 declaration and Vermont acting as the last holdout state and finally declaring a state of emergency on March 16, 2020.[25] The state of emergency declarations, in turn, were shortly followed by stay-at-home orders for most of the 50 states. In addition, the U.S. stimulus check program issued three payments of $600, $2,000, and $1,200 to every resident with an annual income below $60,000. Considering the average Robinhood investor account size is only $3,500, this likely constituted a considerable amount of investable capital for these investors.

As Figure 8 demonstrates, the national and state declarations of emergency coincided with a sharp rise in the number of Robinhood investors. The number of Robinhood investors in the S&P 500 rose sharply starting in spring 2020, going from approximately $40,000$ at the start of the pandemic to nearly $120,000$ by the sample period end date of August 13, 2020. This 300% increase was at least partially in

[24]For the purposes of Figure 8, the Dow Jones stocks were chosen as the constituents of DJIA as of 12/31/2019.

[25]https://en.wikipedia.org/wiki/U.S._state_and_local_government_responses_to_the_COVID-19_pandemic provides a comprehensive timeline of state of emergency declarations and stay-at-home orders in the 50 U.S. states.

response to two specific initiatives that the U.S. employed to combat the pandemic: (1) stay-at-home orders that enforced lockdowns and (2) stimulus checks that increased spending power.

Since two of the three stimulus checks were issued after the end of the Robintrack sample, I focus on the effect of lockdowns.[26] The lockdowns, or state-mandated stay-at-home orders, resulted in thousands of college students being sent home from school and the closure of venues such as restaurants and all but essential stores. This left Americans with more time to spend at home and fewer outside options. As the yellow band designating state lockdowns in Figure 8 demonstrates, this increase in free time coincided with a rise in the popularity of online trading on the Robinhood platform.

Figure 9 demonstrates that the increase in the number of investors following the Covid-19 pandemic was not limited to stocks of a particular size. Instead, the numbers of Robinhood investors in large-cap, mid-cap, and small-cap stocks all increased proportionately at the same growth rate. Panel (a) of Figure 9 plots the median number of Robinhood investors per stock for each of the three market cap groups, and illustrates that all three groups experienced a significant increase during the pandemic. The median number of Robinhood investors in large cap stocks increased the most, appearing to signal that these stocks were disproportionately affected. However, when considering the percent growth of the median number of Robinhood investors for each market cap group over time as plotted in Panel (b) of Figure 9, it becomes evident that all three market size groups actually experienced similar growth rates in the median number of Robinhood investors. The Covid-19 pandemic thus does not appear to have resulted in a disproportionately large growth in demand for any of the three size categories.

Finally, I analyze the growth of the Robinhood platform since the Covid pandemic. Using the two measures of number of funded accounts and net deposits, I demonstrate that the Robinhood platform continued to grow through the end of 2023. As of December 2023, Robinhood had 23.4 million funded accounts on its platform and received an additional $17.1 billion in net deposits. Annual figures for the two growth measures are reported in Appendix G.

[26] On the stimulus check side, Greenwood, Laarits, and Wurgler (2023) employ the method of Boehmer et al. (2021) to estimate stock-day-level measures of retail-initiated buys and sells and find that the first two rounds of stimulus checks in the U.S. appear to have increased retail buying. This evidence is consistent with the increase in Robinhood traders that I find during the Covid-19 pandemic.

3 Differences in Investor Behavior

In this section, I empirically investigate whether Robinhood investors exhibit similar investment behaviors to those observed in previous studies on individual investors. In contrast to prior literature, I find that Robinhood investors do not significantly favor lottery stocks, small-cap stocks, or value stocks. To shed light on this divergence of investing behavior, I employ a regression discontinuity in time design around the introduction of fractional share trading on the Robinhood platform on December 12, 2019. I demonstrate that this market design change helped enable previously capital-constrained investors to buy more high nominal priced stocks, which also tend to be less lottery-like. Finally, I analyze Robinhood investors' reduced interest in small-cap stocks and value stocks and find that it can be explained through extrapolative beliefs in style investing.

3.1 Lottery Stocks

For a long time, individuals have appeared to exhibit a preference for lottery stocks. For instance, Markowitz (1952b) notes that people prefer small chances of large gains with large chances of small losses rather than vice versa.[27] Tversky and Kahneman (1992), Polkovnichenko (2005), and Barberis and Huang (2008) find that investors overweight low probability events and exhibit a preference for stocks with high skewness. Meanwhile, Kumar (2009) establishes a correlation between the propensity for gambling and investment in lottery stocks. Given these established patterns, it is natural to postulate that the new class of retail investors represented by Robinhood traders also invests highly in lottery stocks.

In order to identify lottery stocks, I employ two well-known definitions. The first definition follows Kumar (2009) and identifies lottery stocks as those with the following three characteristics: low price, high idiosyncratic volatility, and high idiosyncratic skewness. Idiosyncratic volatility is calculated as the standard deviation of the residual obtained from fitting a Fama-French-Carhart four factor model to the daily return time series over the previous six month horizon as in Kumar (2009). Idiosyncratic skewness is calculated using the methodology of Harvey and Siddique (2000), who decompose total skewness into idiosyncratic skewness and systematic skewness components. In particular, idiosyncratic

[27]Such behavior cannot be explained by mean-variance preferences as in Markowitz (1952a), but a preference for skewness does emerge in a three moment capital asset pricing model as in Kraus and Litzenberger (1976).

skewness is calculated by taking the third moment of the residual obtained by fitting a two-factor model to the daily stock return time series, where the two factors are the excess market return and squared excess market return. I compute idiosyncratic skewness using daily returns data over the past six months.

At each point in time, I then identify the stocks that fall into the lowest 50th percentile by price, highest 50th percentile by idiosyncratic volatility, and highest 50th percentile by idiosyncratic skewness and denote those as lottery stocks. All three sorts are carried out independently. Conversely, stocks that simultaneously fall into the highest 50th percentile by price, lowest 50th percentile by idiosyncratic volatility, and lowest 50th percentile by idiosyncratic skewness signify non-lottery stocks. All remaining stocks are classified into the other stocks category.

Strictly speaking, the three stock characteristics identify stocks that appear to be like lotteries based on the available information set at a given point in time, rather than stocks that may truly end up being lotteries in the future. One might conceivably wish to classify stocks with a small probability of very large positive returns in the future as lottery stocks using more forward-looking measures. However, while it may be conceivable that sophisticated institutional investors are able to predict future skewness, it is unlikely that less sophisticated individual investors such as those on the Robinhood platform would successfully do so. Instead, they are more likely to naively extrapolate past moments into the future and pick stocks that appear like lotteries based on realized recent volatility and skewness measures as I compute in my first definition of lottery stocks.

The second definition of lottery stocks is based on Barberis and Huang (2008). In this case, a lottery stock is identified by a single characteristic: high idiosyncratic skewness. I compute idiosyncratic skewness following Harvey and Siddique (2000) and use a six month horizon. As mentioned, one advantage to using the two aforementioned definitions of idiosyncratic volatility and idiosyncratic skewness based on realized past returns in the context of retail traders is that these measures are accessible to individual investors. In contrast, more sophisticated measures of volatility such as the forward-looking options implied volatility measure are less likely to be influential for relatively inexperienced and often first-time retail investors.

Table 3 reports the average characteristics of stocks in the lottery stock, non-lottery stock, and other stock categories. Overall, lottery stocks tend to be 51 times smaller by market cap than non-lottery stocks ($402 million versus $20.6 billion). They are also on average two times younger than non-lottery firms, are significantly more illiquid according to the Amihud (2002) illiquidity measure, and are six times less likely to pay dividends. By definition, lottery stocks also have a lower price, higher idiosyncratic volatility, and higher idiosyncratic skewness than their non-lottery stock and other stock counterparts. The average stock price of lottery stocks in the sample is only \$5.71 while that of non-lottery stocks is \$331.42, which signifies an important consideration for capital-constrained investors.

One particularly notable feature of these characteristics is the significantly negative previous year less prior month return of -11.21% exhibited by lottery stocks. The analogous figure for non-lottery stocks is a positive 5.76%. Thus, the fact that Robinhood investors did not overinvest in lottery stocks aided their performance over the period. The second takeaway of the lottery stock versus non-lottery stock return differential is that if individual investors in the Robinhood sample have extrapolative beliefs, it would be natural for them to move away from investing in lottery stocks following those stocks' recent underperformance. As I show in the value stocks and small-cap stocks subsection, the absence of investment in small-cap stocks and value stocks is also consistent with investors extrapolating prior returns for these investment styles.

Figure 10 plots a time series of the average number of Robinhood investors in lottery stocks and in non-lottery stocks over time. Although there is a small uptick in the average number of investors in lottery stocks near the end of the sample period, for the majority of the time period the average number of investors in lottery stocks and non-lottery stocks was roughly equal. This supports the notion that for the most of the sample period, Robinhood investors did not exhibit a preference towards lottery stocks. Figure 11 corroborates this finding and shows the total number of Robinhood investors aggregated across all lottery stocks, non-lottery stocks, and other stocks. The other stocks category remains the dominant category over the sample period, and experiences a high level of growth. In a robustness analysis in Figure IA.3, I also plot the aggregate and mean number of Robinhood investors in lottery, non-lottery, and other stocks when employing an alternative definition of lottery stocks with percentile parameter $k = 40$. In this case, lottery stocks are categorized as those in the lowest 40% by price, highest 40% by idiosyncratic volatility, and highest 40% by idiosyncratic skewness. Similarly, in the robustness analysis shown in Figure IA.4 I employ the definition of lottery stocks based on three independent sorts on the basis of the percentile parameter $k = 30$. The mean numbers of Robinhood investors per lottery stock and per other stock were roughly equal over the sample period.

In order to test for investment in lottery stocks more formally, I perform two sets of analyses. The first is a set of panel regressions where the dependent variable is a scaled measure of the number of Robinhood investors and the independent variables of interest include the three lottery stock characteristics. I also include size and value regressors, where size is measured as the natural log of market cap and value is measured as the book-to-market ratio. The regression includes time fixed effects, which is important since the Robinhood platform experienced a significant rise in popularity over the sample period. Standard errors are clustered by time and security, and the control variables include momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, dividends per share, an indicator variable for a dividend-paying firm, firm age, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. The independent variables of interest are standardized. More detailed variable descriptions and computation methodology are provided in Appendix A.

The full regression specification is provided below:

$$\log(NI_{i,t+1}) = \beta_1 Prc_{i,t} + \beta_2 IVol_{i,t} + \beta_3 ISkew_{i,t} + \beta_4 \log(ME)_{i,t} + \beta_5 BM_{i,t} + \gamma' Controls_{i,t} + \delta_{t+1} + \epsilon_{i,t+1} \tag{1}$$

where $Prc_{i,t}$ is the dollar price of stock i on date t, $IVol_{i,t}$ is the idiosyncratic volatility of stock i computed using returns data ending on date t, $ISkew_{i,t}$ is the idiosyncratic skewness of stock i computed using returns data ending on date t, $\log(ME)$ is the log market cap of stock i on date t, BM is the winsorized book-to-market ratio of stock i on date t, and $Controls$ are the control variables outlined above. The regression model includes time fixed effects δ_{t+1}. The dependent variable $\log(NI_{i,t+1})$ is a cleaned and scaled version of the number of Robinhood investors in stock i on date $t+1$. Specifically, I perform the following three steps to arrive at the final dependent variable $\log(NI_{i,t+1})$:

1. I first limit observations to those with a positive number of Robinhood investors. The rationale for this is to limit instances where Robintrack data became available at a sudden point in time, and is backfilled with zeros prior to a sudden large change.

2. I next winsorize the raw number of investors in stock i on date $t+1$ at the 0.5% and 99.5% level in order to limit the impact of outliers.

3. Finally, I take the log of the winsorized value to scale the dependent variable.

I estimate equation (1) and report the results in Table 4. Column (1) shows the results of a panel regression that includes the three lottery stock characteristics of previous day stock price, idiosyncratic volatility, and idiosyncratic skewness as well as time fixed effects. The coefficient on price is positive and the coefficient on idiosyncratic volatility is positive and statistically significant at the 1% level, while the coefficient on skewness is negative and statistically significant at the 1% level. Thus out of the three lottery stock characteristics, only the coefficient on volatility goes in the same direction as a preference for lottery stocks would suggest; the coefficients on price and skewness go the opposite direction, indicating that Robinhood investors do not exhibit a preference for investing in lottery stocks. In terms of magnitude, a one standard deviation increase in stock price corresponds to an $e^{4.60} - 1 \approx 98.5\%$ increase in the number of Robinhood investors in a stock. Analogously, a one standard deviation increase in idiosyncratic volatility corresponds to an $e^{0.54} - 1 \approx 0.72\%$ increase in Robinhood investors and a one standard deviation increase in idiosyncratic skewness corresponds to an $e^{-0.13} - 1 \approx -0.12\%$ decrease in Robinhood investors. The magnitude of the stock price coefficient points to the preeminence of price as a key consideration for Robinhood traders.

The results of the full control specification are summarized in column (4). The coefficients on price, idiosyncratic volatility, and idiosyncratic skewness maintain the same sign as earlier. Once again, only one of the three characteristics—idiosyncratic volatility—goes in the direction suggested by lottery stock preference. Consequently, I conclude that Robinhood investors do not exhibit a preference for lottery stocks, a surprising finding in contrast to studies of individual investors in the past.

One consequence of the above is that Robinhood investors also do not exhibit a preference for idiosyncratic skewness. This indicates that employing the second definition of lottery stocks as those with high idiosyncratic skewness according to Barberis and Huang (2008) yields consistent results. Namely, the negative and significant coefficient on idiosyncratic skewness illustrates that Robinhood investors do not exhibit a lottery stock preference according to this definition of lottery stocks either.

The second set of analyses I perform are Fama and MacBeth (1973) regressions, the results of which are presented in column (5) and column (6) of Table 4. The direction and magnitude of these results closely mirror the panel regression results in columns (1) and (4) respectively. The consistency between both methods of estimating the impact of lottery stock characteristics on future Robinhood investment lends credence to the observed lack of lottery stock investing behavior.

In additional robustness exercises, I compute panel regression and Fama and MacBeth (1973) regression estimates for each of the calendar years 2018, 2019, and 2020 separately. The results are presented in Table IA.9, Table IA.10, and Table IA.11 respectively of Appendix D and provide evidence consistent with the main analysis results, namely the absence of lottery stock investing by Robinhood traders. In a second type of robustness test, I employ an alternative dependent variable equal to the percentile of the number of Robinhood investors and find that using the percentile measure also yields similar results. The full results for the percentile specification are presented in Table IA.12 for the full sample period and in IA.13 for each of the calendar years separately. In a third exercise, I employ total volatility and total skewness measures in place of idiosyncratic volatility and idiosyncratic skewness.[28] The results yield estimates consistent with the main takeaway, and are presented in Table IA.14 of Appendix D. Finally, in a fourth test I include industry fixed effects in addition to time fixed effects; the coefficients do not change significantly from the main specification, and are presented in Table IA.16.

3.2 Introduction of Fractional Share Trading

On December 12, 2019, Robinhood introduced fractional share trading on its platform. This institutional design change affected previously capital-constrained investors, and increased their ability to purchase high-priced stocks. As a result, I hypothesize that this design change resulted in lower retail ownership of lottery stocks, which tend to be low-priced.

In order to test whether the introduction of fractional share trading impacted Robinhood investment in lottery stocks, I employ the following regression discontinuity in time (RDiT) design around the introduction date of December 12, 2019. Firm observations are "treated" (able to be traded in fractional amounts) based on a known cut-off rule (time $t > t'$, where t' is the date fractional share trading was introduced). Time t is the forcing variable, and December 12, 2019 is the threshold value t'. $p(0)$ is the outcome absent treatment, and $p(1)$ is the outcome with treatment. I am interested in how this treatment affects the outcome variable of interest, the loading on price p_t.

I then estimate the following regression specification:

[28] Total skewness is referred to as a measure of lottery stock investing in Garrett and Sobel (1999), and Kumar (2009) uses total skewness as a robustness measure.

$$p_t = \alpha + \beta Post_t + \gamma^b(t - t') + \gamma^a Post_t(t - t') + u_i \tag{2}$$

where p_t is the loading on price of Robinhood investment, t is time, t' is the date fractional share trading was introduced, and $Post_t$ is an indicator variable for whether the date occurs after this introduction.

I confirm that the following assumptions hold:

- **RDiT Randomization Assumption** — Firms cannot manipulate time, so whether an observation's date falls immediately above or below the fractional share trading introduction date is random.

- **Sharp RDiT Assumption** — Assignment to treatment occurs through a known and deterministic decision rule:

$$d = d(t) = \begin{cases} 1 & \text{if } t \geq t' \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

- **RDiT Local Continuity Assumption** — The potential outcomes, $p(0)$ and $p(1)$, conditional on forcing variable t, are continuous at threshold t' (i.e., p would be a smooth function around December 12, 2019 absent treatment of the fractional share introduction change—there was no other platform innovation around this time).

I find that the magnitude of the discontinuity is $\beta = 1.43$ with a t-stat of 4. This indicates a positive and significant change in price loading following the introduction of fractional share trading, and corresponds to a 32% increase relative to the average price loading over the sample period. These results demonstrate the connection between the adoption of fractional share trading and the decline in investment in lottery stocks by the emerging retail investor class.

The implications of these results are two-fold. First, they point towards the importance of considering market frictions, innovations, and constraints when analyzing investor behavior. Portfolio holdings that may at first appear to signify investor preferences may instead be a product of frictions and constraints imposed on investment. Specifically, in the case of lottery stocks, the loosening of capital constraints appears to have allowed investors to move away from lottery stock-heavy U.S. stock portfolios.

Second, these findings highlight the dynamic nature of stock category definitions, emphasizing the need for periodic reassessment in academic research. Particularly, the reliance on low price as a defining characteristic of lottery stocks may be becoming obsolete in a landscape where even the smallest retail investors possess the ability to invest in some of the U.S. stock market's most highly-priced stocks with minimal amounts, sometimes as low as one dollar.

3.3 Value Stocks and Small-Cap Stocks

I also document that in contrast to previously studied individual investors, Robinhood investors are not highly invested in value stocks and small-cap stocks. In fact, Robinhood investors are significantly overweight in large-cap stocks and growth stocks.

Column (3) of Table 4 investigates whether Robinhood investors have a preference for value. I find that the coefficient on the book-to-market ratio is negative and statistically significant at the 1% level, and remains so when I add all of the control variables as shown in Column (4). This indicates that unlike previously studied individual investors, Robinhood investors do not exhibit a preference for value stocks.

Column (2) of Table 4 regresses the number of Robinhood investors on previous day size, where size is measured as the log market cap. The coefficient is positive and significant at the 1% level, indicating that all else equal Robinhood investors prefer large-cap rather than small-cap stocks. The coefficient on size remains positive and statistically significant at the 10% level even after including all of the control variables. One possible explanation for this could be that Robinhood introduced fractional share trading in December 2019, at which point investors could more easily buy stocks with high share prices. Stocks with high share prices, in turn, are correlated with stocks that have a large market cap. Figure 12 investigates this explanation via a series of monthly panel regressions of the number of Robinhood investors on previous day stock price, idiosyncratic volatility, and idiosyncratic skewness. The coefficients on price from these monthly regressions are plotted in the figure, and demonstrate a relative rise in 2020 following the introduction of fractional share trading on the Robinhood platform in December 2019.

Another possible contributing explanation comes from the common assumption in psychology-based models of asset prices and investor behavior that people have extrapolative demand. Specifically, an individual's demand for a financial asset can depend positively on the asset's past returns, and especially on its recent past return.[29] To test this theory and determine whether the observed demand for growth and large-cap stocks is consistent with extrapolative demand, I compute the returns on the value (HML) and size (SMB) factors in the years preceding the Robinhood sample and the Barber and Odean (2000) sample. The results are reported in Table 5. I find a significant difference in returns, namely that returns of the value and size factors were significantly lower in the years leading up to the Robinhood sample years. This is supportive evidence for extrapolative return beliefs in style investing among retail traders.

4 Three Key Determinants of Robinhood Investment

In this section, I identify three key determinants of Robinhood investment. I first introduce the novel buy-the-dip effect, and show how Robinhood investment for large-cap companies is highest following extreme negative returns and earnings surprises. I propose a model based on the gambler's fallacy to explain the buy-the-dip effect. I next demonstrate evidence of event-based trading by Robinhood traders in response to earnings announcements and analyst recommendation revisions. Finally, I show that peer effects based on both popularity and sentiment on the WallStreetBets platform predict Robinhood investment and propose a novel sentiment dictionary that incorporates slang terminology and graphical representations of emojis.

4.1 Buy-the-Dip Effect

In this subsection, I document a novel investing preference of the Robinhood investors, which I call the buy-the-dip effect. The essence of the buy-the-dip effect is as follows: Robinhood investors invest most in large-cap stocks following extreme negative events. This is not the case for small-cap stocks, in which Robinhood investment is highest following extreme positive events. I provide evidence of the buy-the-dip effect in Robinhood top holdings, through quintile analysis, and via regression analysis.

[29] For a selection of both theoretical and empirical work related to extrapolation, see Barberis and Shleifer (2003), Barberis et al. (2015), Barberis et al. (2018), Bastianello and Fontanier (2022), Cutler, Poterba, and Summers (1990), De Long et al. (1990), Greenwood and Shleifer (2014), and Liao, Peng, and Zhu (2022).

4.1.1 Buy-the-Dip Effect in Top Holdings

As most other retail investors, Robinhood investors exhibit some degree of brand affinity towards well-known and well-liked companies, which tend to be larger in size. While Robinhood investors would like to invest in these high brand-affinity, well-liked companies, they may feel hesitant to do so because they perceive that the price of the company is already "too high." In fact, financial advisors must frequently overcome this obstacle with their clients to encourage them to invest in the stock market as soon as possible, rather than waiting until it dips and a so-called "better opportunity" arises. As a result, after an extreme negative return in a high brand affinity company, Robinhood investors are more likely to feel good about their decision to purchase it and this leads to the buy-the-dip effect.

Table 1 lists the top 10 Robinhood U.S. stock holdings at the end of the sample period on August 13, 2020. The list includes two large, well-known, troubled automakers and two large, well-known airline companies that faced especially tough market conditions and performance during the pandemic. Robinhood investment in these large, well-known companies that fell upon hard times is consistent with the buy-the-dip effect.

Figure 13 delves further into the performance of each of the top 10 Robinhood holdings. Notably, six of the top 10 holdings experienced a drawdown of more than 50% since the beginning of the sample period, consistent with the buy-the-dip effect. Furthermore, the WallStreetBets forum frequently includes bullish posts that include the phrase "buy the dip" and talk about investing in a well-liked, popular company after its price has experienced a drop. Although my Robinhood sample ends before the rally in Gamestop (GME) in January 2021 and the rally in AMC Entertainment (AMC) over May and June 2021, these investments would also be consistent with the buy-the-dip effect where Robinhood investors buy well-liked companies after they fall upon hard times.

4.1.2 Buy-the-Dip Effect by CAR Quintile

I next demonstrate the buy-the-dip effect using data for all Robinhood holdings rather than a sub-sample limited to the top holdings. In particular, I consider three-day percent changes in the number of Robinhood investors for companies in five quintiles, sorted according to their contemporaneous three-day cumulative abnormal return (CAR). CAR is computed using compounded company returns net of the risk-free rate, and CARs are winsorized at the 0.5 and 99.5 levels prior to sorting into quintiles. In a robustness check, I also compute CAR using CAPM alphas and find similar results.[30]

Table 6 shows the average three-day percent change in the number of Robinhood investors for stocks in a given CAR quintile. Quintile 1 represents the lowest CAR quintile and Quintile 5 represents the highest CAR quintile. Small cap includes all observations for stocks with a market cap less than or equal to $2bn, mid cap includes all observations for stocks with a market cap between $2bn and $10bn, and large cap includes all observations for stocks with a market cap over $10bn. The Overall column includes observations across all market caps, and the Observation Count column reports the number of observations used to compute the quintile averages in the Overall column.

Consistent with my results on Robinhood investors and event-based trading, the number of Robinhood investors increases following an extreme return. The Overall column in Table 6, which includes all observations across all market caps, shows that while the average increase in the number of Robinhood investors in a stock whose CAR falls into the medium Quintile 3 was just 0.4%, the analogous average percent increases in the number of Robinhood investors in the highest Quintile 5 and lowest Quintile 1 stocks were significantly higher at 2.4% and 1.8%, respectively.

However, there is also a pronounced difference between small-cap and large-cap stocks which represents the buy-the-dip effect. For large-cap stocks, Robinhood investors invest the most when the return falls into the lowest quintile, as can be seen from the Large Cap column in Table 6. This is consistent with the buy-the-dip effect for large-cap stocks. For small-cap stocks, however, the highest percent increase in the number of Robinhood investors occurs when the return is in the highest quintile. Thus, the buy-the-dip effect is only present for large-cap companies.

4.1.3 Buy-the-Dip Effect Regression Analysis

While the preceding subsection provides illustrative evidence for the buy-the-dip effect, the quintile double sort analysis includes contemporaneous changes in CAR and in the number of Robinhood investors. In this subsection, I present the regression results of the percent change in the number of Robinhood investors on the previous CAR quintile in order to better identify the effect that CAR quintile has on future Robinhood investment.

[30] The results for the robustness analysis employing CAPM alphas are reported in Appendix D, Table IA.17.

The full regression specification is as follows:

$$\Delta \log NI_{i,t+1} = \beta_1 \mathbb{1}_{i,t}^{Q1or5} + \beta_2 \mathbb{1}_{i,t}^{Q5} + \sum_{\tau=t-4}^{t} \beta_{3,\tau} \Delta \log NI_{i,\tau} + \gamma' Controls_{i,t} + \delta_{t+1} + \epsilon_{i,t+1} \qquad (4)$$

where $\Delta \log NI_{i,t+1}$ is the change in the log winsorized number of Robinhood investors in stock i from date t to date $t+1$ (note that this is approximately equal to the daily percent change in Robinhood investors, since $\Delta \log NI_{i,t+1} = \log NI_{i,t+1} - \log NI_{i,t} = \log \frac{NI_{i,t+1}}{NI_{i,t}} = \log(1+\frac{NI_{i,t+1}-NI_{i,t}}{NI_{i,t}}) = \log(1+\%\Delta NI_{i,t+1}) \approx \%\Delta NI_{i,t+1}$ as $\log(1+x) \approx x$ for small values of x). In order to reduce the impact of outlier percent changes, I limit the sample to observations where the number of Robinhood investors in a stock is greater than 10. The new variable $\mathbb{1}_{i,t}^{Q1or5}$ is an indicator variable equal to 1 if the previous period return of stock i was in either the extreme positive or extreme negative quintile and equal to 0 otherwise, and $\mathbb{1}_{i,t}^{Q5}$ is an indicator variable equal to 1 if the previous period return of stock i fell into the highest return quintile and equal to 0 otherwise. The remaining variables hold the same meaning as in the prior regression specification (1).

I estimate equation (4) separately for the sample of large-cap stocks and the sample of small-cap stocks. The results are presented in Table 7. For both the large-cap and the small-cap stock samples, the coefficient on the indicator variable $\mathbb{1}^{Q1or5}$ is positive and statistically significant at the 1% level. This indicates that Robinhood investors invest more following both extreme positive and extreme negative returns, consistent with my findings on event-based trading in Section 4.2. However, for small-cap stocks the coefficient on $\mathbb{1}^{Q5}$ is positive and statistically significant at the 1% level, while for large-cap stocks the same coefficient is negative and statistically significant at the 1% level. This indicates that there is a buy-the-dip effect for large-cap stocks, and Robinhood investors invest relatively more in large cap stocks following extreme negative returns. For small-cap stocks, however, I do not observe the buy-the-dip effect as Robinhood investors invest relatively more following extreme positive returns.

Finally, I also analyze a single regression specification that includes all observations among all market cap sizes. Among the regressors, I include both the previous day return and the previous day absolute value of return. In order to see the differential effect between small-cap and large-cap investments, I also include two new variables among the regressors.

The first new variable is $RetSC_{i,t}$, which is the previous day return of stock i multiplied by an indicator variable for whether stock i is a small-cap stock. The second new variable is $RetLC_{i,t}$, which is the previous day return of stock i multiplied by an indicator variable for whether stock i is a large-cap stock. Note that I do not include an analogous variable for mid-cap stocks in order to avoid multicollinearity of the regressors. The regression includes control variables and time fixed effects, and standard errors are clustered by time and security. The full regression specification is:

$$\begin{aligned} \Delta \log NI_{i,t+1} = {} & \beta_1 Ret_{i,t} + \beta_2 RetAbs_{i,t} + \beta_3 RetLC_{i,t} + \beta_4 RetSC_{i,t} \\ & + \sum_{\tau=t-4}^{t} \beta_{5,\tau} \Delta \log NI_{i,\tau} + \gamma' Controls_{i,t} + \delta_{t+1} + \epsilon_{i,t+1} \end{aligned} \tag{5}$$

Table 8 presents the results. Consistent with my event-based trading findings, I find a positive and statistically significant coefficient on the previous day absolute value of return, indicating that Robinhood investors invest more following extreme returns. I also find a negative and statistically significant coefficient on the previous day return, indicating that the Robinhood investors may be more contrarian overall. Importantly for the buy-the-dip effect, I also document a difference between the large-cap and small-cap return coefficients, $RetLC_{i,t}$ and $RetSC_{i,t}$. This difference is statistically significant and provides further evidence for the buy-the-dip effect in large-cap stocks.

In particular, the coefficient $RetLC_{i,t}$ is negative and statistically significant, indicating that Robinhood investors invest relatively more in large-cap companies with extreme negative returns. For small-cap companies I observe a different result: the coefficient $RetSC_{i,t}$ is positive and statistically significant, indicating that Robinhood investors invest relatively more in companies with positive returns when those companies are small cap.

4.1.4 Model

I next propose a simple belief-based model that can help explain the observed novel buy-the-dip phenomenon. In particular, the unifying explanation for this rests on individual investors' misplaced and well-documented belief in the law of small numbers (Kahneman and Tversky (1971), Rabin (2002), Rabin and Vayanos (2010)). The law of small numbers posits that individuals tend to expect to see large sample properties in small samples. For instance, individuals tend to underestimate the frequency of extreme outcomes in a short sequence of coin tosses and expect to see a more equal distribution.

One behavioral bias based on the law of small numbers is the gambler's fallacy. The gambler's fallacy states that following a sequence of the same type of outcome, individuals expect that it is time for a different type of outcome to occur. This is a fallacy if the outcomes are in fact independent of each other. After a large-cap stock has experienced a negative event, or a dip, individuals suffering from the gambler's fallacy would place a higher likelihood on a different outcome (in this case, a stock price rise) in the following period. In the context of large-cap stocks, the buy-the-dip effect is thus consistent with investors suffering from the gambler's fallacy.

Model Setup. Following Rabin (2002), an individual observes a sequence of binary signals $\{s_t\}_{t=1}^{\infty}$ that are generated randomly from a stationary probability rate $\theta \in [0, 1]$. The binary signals represent performance signals with $s_t \in \{h, l\}$ and $P[s_t = h] = \theta$. Signal h denotes a high value and signal l denotes a low value on any given day. In the context of stock return expectations, these high and low values could denote (i) the returns themselves, (ii) a parameter for the mean of a normal return distribution, or (iii) the parameter for the mean of a lognormal return distribution depending on the asset pricing framework considered. Let Θ denote the set of rates that occur with positive probability, so that the rate θ occurs with prior probability $\pi(\theta) > 0$ and $\sum_{\Theta} \pi(\theta) = 1$.

The individual who observes the performance signals is a Bayesian and has correct priors about the rate θ. However, whereas signals are independent and identically distributed in reality, the individual mistakenly believes that they are generated by random draws without replacement from a sample of size $N < \infty$. This assumption effectively captures the law of small numbers, since it implies that the individual believes the proportion of signals must "balance out" to the population rate before N signals are observed. As $N \to \infty$, the individual becomes fully Bayesian while the smaller the N, the more the individual believes in the law of small numbers.

For a given positive integer N and rate θ, the individual believes the performance signals are drawn without replacement from a sample with exactly θN high signals and $(1 - \theta)N$ low signals. When observing a long sequence of signals, the individual believes that the sample is renewed after every two draws. Two further assumptions ensure that the individual believes all sequences of signals are possible:

1. For all possible rates $\theta \in \Theta$, θN is an integer

2. There exists a possible rate $\theta \in \Theta$ such that $\min[\theta N, (1 - \theta)N] \geq 2$

The model leads to the gambler's fallacy, as due to the nature of the limited draws in the sample the individual expects the second draw of a signal to be negatively correlated with the first draw.

Gambler's Fallacy Lemma. Consider sample size N and rate θ such that θN is an integer and $\pi(\theta) = 1$. Let $\pi_t^N(s_t = s)$ denote the individual's probability that the time t signal will be s. For all even $t \geq 2$ and histories of signal draws h_{t-2}:

$$\pi_t^N(s_t = h | s_{t-1} = l, h_{t-2}) = \frac{\theta N}{N-1} > \theta \tag{6}$$

$$\pi_t^N(s_t = h | s_{t-1} = h, h_{t-2}) = \frac{\theta N - 1}{N-1} < \theta \tag{7}$$

The lemma indicates that consistent with the gambler's fallacy, following a signal of a certain type the individual would perceive a lower probability of the same signal being repeated in the next period and a higher probability of a different signal being repeated in the next period and increase his probability of a different signal appearing in the next period. The intuition in the case of large-cap stocks and buy-the-dip behavior is that individuals are more likely to have strong prior beliefs about the expected returns for large, well-known companies. Consequently, they interpret more of the observed innovations in realized return as noise rather than actual changes in the underlying state variable of expected returns, and expect mean reversion in the next period.

Conversely, individuals are less likely to have a strong belief about expected returns for smaller, less well-known companies. Consequently, they interpret more of the observed innovations in realized return as an underlying expected return change rather than shocks, luck or noise. They therefore update their future expected returns distributions contingent on this new signal, and do not expect mean reversion as in the large-cap case.

I provide a direct test of this underlying logic in column (3) of Table 8. Specifically, I measure the buy-the-dip effect in the 100 most popular stocks in Robintrack and compare it to the buy-the-dip effect in other stocks. If the underlying logic is true—i.e., investors have stronger prior beliefs about large-cap, well-known companies and that is what is driving the buy-the-dip effect—I would expect the buy-the-dip effect to be stronger in the most popular stocks as those should be the stocks which the Robinhood investors have been watching and are likely to have the strongest priors about. Indeed, I find this to be the case.

Example. As an example, let $N = 4$ and $\theta = 0.5$. Suppose Maria believes that a large-cap Company A will outperform with probability 1/2. If $N = 4$, Maria believes Company A has two high return days and two low return days coming. This corresponds to parameter values $\Theta = \{1/2\}$,

$\theta = 1/2$ with probability 1. $N = 4$ so the sample contains $\theta N = (1/2)4 = 2$ high return day signals and $(1-\theta)N = (1/2)4 = 2$ low return day signals.

Let us now assume that Company A experiences a dip, i.e. a low return day. After one low return day is used up, Maria's beliefs (the "sample" she has in mind) contain one more low return day signal and two more high return day signals. If Company A underperforms in its first day, Maria thus believes there is only a 1/3 chance of underperformance the following day (since one of the bad days has already been "used up"). Therefore, Maria is likely to "buy the dip" for Company A.

Portfolio Choice. I next incorporate the gambler's fallacy belief formation process into a constant relative risk aversion (CRRA) and lognormal returns framework in the following manner. Consider an individual who has power utility with relative risk aversion γ. The individual has the option to invest in the riskless asset with return R_f and in the risky asset with return R_t, where the mean of the risky asset's return is based on the individual's beliefs about the remaining performance signals in the sample. Specifically, the individual's expectations for the risky asset return are given by:

$$\mathrm{E}_{t-1}[\mu_t] = \begin{cases} \left(\frac{\theta N}{N-1}\right)h + \left(\frac{N-1-\theta N}{N-1}\right)l & \text{if } s_{t-1} = l \\ \left(\frac{N-\theta N}{N-1}\right)l + \left(\frac{\theta N-1}{N-1}\right)h & \text{if } s_{t-1} = h \end{cases} \tag{8}$$

The individual chooses to invest portfolio share α_t in the risky asset, and the individual's optimization problem is:

$$\max_{\alpha_t} \mathrm{E}\Big[\frac{W_{t+1}^{1-\gamma} - 1}{1-\gamma}\Big] \tag{9}$$

subject to the budget constraint $W_{t+1} = (1 + R_{p,t+1})W_t$. The budget constraint can be rewritten using log returns and log wealth as $w_{t+1} = r_{p,t+1} + w_t$.

Using a second-order Taylor approximation, the log portfolio return $r_{p,t+1}$ can be written as follows:

$$r_{p,t+1} = r_{f,t+1} + \alpha_t(r_{t+1} - r_{f,t+1}) + \frac{1}{2}\alpha_t(1-\alpha_t)\sigma_t^2 \tag{10}$$

The optimal level invested in the risky asset is then given by:

$$\alpha_t^* = \frac{\mathrm{E}_t[R_{t+1}] - R_{f,t+1}}{\gamma\sigma_t^2} = \begin{cases} \frac{\left(\frac{\theta N}{N-1}\right)h + \left(\frac{N-1-\theta N}{N-1}\right)l - R_{f,t+1}}{\gamma\sigma_t^2} & \text{if } s_t = l \\ \frac{\left(\frac{N-\theta N}{N-1}\right)l + \left(\frac{\theta N-1}{N-1}\right)h - R_{f,t+1}}{\gamma\sigma_t^2} & \text{if } s_t = h \end{cases} \tag{11}$$

Therefore, the individual invests more in the risky asset following a dip in the prior period, and less following a positive prior period performance signal. This is consistent with the buy-the-dip effect for large-cap stocks. Additionally, the results hold in alternative specifications such as the CARA-Normal setting or in a discrete-time gamble setting. Adding the limited attention component to the model (i.e., restricting the available choice set to assets that attract investor attention) enables it to explain both the buy-the-dip effect (through the gambler's fallacy component outlined here) and event-based trading and WallStreetBets (through the limited investment choice set component). In this way, a parsimonious model is able to explain all of the observed empirical key drivers for Robinhood investment.

Alternative Explanations. I am able to exclude several alternative explanations for the buy-the-dip effect, which I briefly summarize below.

Firstly, the buy-the-dip effect is not consistent with rational explanations since following the release of informative negative news, the stock price of a company is expected to decline and investors are expected to sell. This is the opposite from what occurs for large-cap companies under the buy-the-dip effect, where investors buy following extreme negative news, and is thus not a consistent explanation for this behavior.

Second, several behavioral finance explanations also fail to explain the buy-the-dip effect. For instance, an investor with extrapolative beliefs—following the observation of a negative return—would expect the negative returns to continue. Consequently, an extrapolator would not enter into a position to buy the dip of a large-cap company that has experienced a recent negative return. Other behavioral finance explanations based on the notions of probability weighting and loss aversion also fail to explain the distinct behavior of Robinhood investors towards large-cap versus small-cap stocks.

Third, I consider an explanation based on mean-variance portfolio rebalancing. Under this supposition, an individual would behave in an observed contrarian manner in order to maintain their chosen portfolio allocation to various asset classes and security types.[31] For instance, following positive returns an individual may wish to sell the security in order to decrease the portfolio weight of the security

[31] A more formal explanation of mean-variance portfolio rebalancing is provided in Appendix F.

back to the predetermined optimal level. Similarly, an individual may buy the dip for a security which has suffered a loss to bring its portfolio weight back up. However, this mechanism fails to explain the observed differential trading behavior of Robinhood investors between large-cap and small-cap stocks. In particular, I find that Robinhood investors buy the dip only for large-cap stocks; therefore, the mean-variance portfolio rebalancing explanation is insufficient to explain their behavior.

Fourth and finally, I analyze an explanation for the buy-the-dip effect that is based on tax loss harvesting. Badrinath and Lewellen (1991), Odean (1998) and Grinblatt and Keloharju (2001) find evidence that tax losses tend to be realized at the end of the year, particularly in the second half of December, consistent with the tendency for large losses to be realized at the last minute. If the buy-the-dip behavior only occurred at year-end and only small-cap stocks had suffered losses over the year, tax loss harvesting could provide a plausible explanation. However, on the contrary I document that the buy-the-dip effect is not limited to December but rather occurs throughout the sample period, including over the January 2020 - August 2020 period.[32] Therefore, tax loss harvesting does not provide an explanation for the buy-the-dip effect.

4.2 Event-Based Trading

The second main determinant of Robinhood investment I identify is event-based trading, which includes trading in response to earnings announcements, analyst recommendation revisions, and extreme returns, volatility, and volume.

4.2.1 Extreme Returns, Volatility, and Volume Traded

I first test the influence of previous day extreme returns, volatility, and volume traded on Robinhood next day investments. Due to the short horizon of the one-day returns, I employ a regression using the percent changes of the number of Robinhood investors in stock i at time t as the dependent variable. I include time fixed effects in the regression, and standard errors are clustered by time and security. The

[32] In addition, when I break out the sample period into both odd and even months, or into separate calendar years, I find the results on the buy-the-dip effect are largely unchanged.

full regression specification is as follows:

$$\Delta \log NI_{i,t+1} = \beta_1 Ret_{i,t} + \beta_2 RetAbs_{i,t} + \beta_3 IVol_{i,t} + \beta_4 DolVolume_{i,t} \\ +\beta_5 AbnDolVolume_{i,t} + \sum_{\tau=t-4}^{t} \beta_{6,\tau} \Delta \log NI_{i,\tau} + \gamma' Controls_{i,t} + \delta_{t+1} + \epsilon_{i,t+1} \tag{12}$$

where the dependent variable $\Delta \log NI_{i,t+1}$ is the one-day percent change in the number of Robinhood investors in stock i from day t to day $t+1$, $Ret_{i,t}$ is the previous day return of stock i, $RetAbs_{i,t}$ is the absolute value of the previous day return of stock i, $IVol_{i,t}$ is the idiosyncratic volatility of stock i, $DolVolume_{i,t}$ is the previous day dollar volume traded of stock i (computed as previous day share volume times previous day price, both taken from CRSP), and $AbnDolVolume_{i,t}$ is the abnormal dollar volume traded. The $AbnDolVolume_{i,t}$ variable is computed following Barber and Odean (2008) as:

$$AbnDolVolume_{i,t} = \frac{DolVolume_{i,t}}{DolVolumeAve_{i,t}} \tag{13}$$

where $DolVolume_{i,t}$ is the dollar volume traded for stock i on date t and $DolVolumeAve_{i,t}$ is the average of the dollar volume traded for stock i in the preceding 252 days (days $t-1$ to $t-252$).

Finally, the regression also includes the five previous lagged percent changes in the number of Robinhood investors in stock i, time fixed effects, and control variables such as momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, share and dollar volume traded, Amihud illiquidity, and bid-ask spread.

Table 9 presents the result. Column (1) shows the results for a regression of the percent change in the number of Robinhood investors on previous day return and previous day absolute value of return. The coefficient on the previous day absolute value of return is positive and statistically significant at the 1% level, indicating that Robinhood investors demonstrate a predilection for buying after extreme returns. The coefficient on the previous day return, meanwhile, is negative and statistically significant, indicating that Robinhood investors overall tend to be more contrarian. The coefficients on the previous day return and previous day absolute value of return maintain their sign and significance at the 1% level in the full specification with controls in Column (3) as well, indicating the robustness of the results.

Column (2) reports the results with abnormal volume traded and total previous day volume traded as the independent variables. Neither coefficient is statistically significant, though both are positive. In the full specification in Column (3), both coefficients are positive and the coefficient on previous day dollar volume traded is statistically significant at the 1% level. This result is consistent with Welch (2022), who finds that trading volume significantly helps explain holdings in the aggregate Robinhood portfolio. It also indicates that Robinhood investors may prefer to trade liquid and newsworthy stocks that are experiencing high volume traded.

4.2.2 Earnings Announcements

In order to further test whether Robinhood traders engage in event-based trading, I study how Robinhood investors respond to earnings announcements and earnings surprises. In this section I consider a three-day short-window event study around earnings announcement days, where the average abnormal increase in Robinhood investors is used as a gauge of how they respond to the earnings surprises. I use a three-day window in order to capture all of the reaction around the earnings surprise.

Standardized unexpected earnings (SUE) are computed as the reported earnings per share less analyst forecast earnings per share in the same quarter, divided by the standard deviation of the analyst forecasts in that quarter. I sort the standardized unexpected earnings over the sample period into five quintiles, with Quintile 1 being the lowest and Quintile 5 being the highest. I then compute the contemporaneous three-day percent change in the number of Robinhood investors for each earnings surprise, and record the averages by SUE quintile and market cap size in Table 10. The control group includes all observations that did not have an earnings announcement. Small cap includes all observations for stocks with a market cap less than or equal to $2bn, mid cap includes all observations for stocks with a market cap between $2bn and $10bn, and large cap includes all observations for stocks with a market cap over $10bn. The Overall column includes observations across all market caps.

The results indicate two main findings. The first is that Robinhood investors increase their holdings in stocks that report earnings, relative to stocks with no earnings announcements. The second is that this increase in holdings is higher for stocks whose earnings surprises are extreme.

The first finding can be seen by comparing the bottom Control row, which averages across observations with no earnings announcements, in Table 10 with the rows above it. On average in the control group, stocks experience a 1.1% increase in the number of Robinhood investors over a three-day period. The analogous three-day percent increases for the overall sample for stocks that had earnings announcements ranges from 5.0% to 6.6%, depending on the SUE quintile. Therefore Robinhood investors invest considerably more in stocks that have earnings announcements, which is consistent with event-based trading.

The second finding can be illustrated by comparing the Quintile 1 (extremely negative earnings surprise) and Quintile 5 (extremely positive earnings surprise) rows with the Quintile 3 row (neither extremely negative nor extremely positive earnings surprise). Across every market cap group as well as the Overall group, the number of Robinhood investors increases more for earnings announcements that fall into the extreme positive and extreme negative Quintiles 5 and 1 than for earnings announcements that fall into the middle Quintile 3. This indicates that Robinhood investors invest relatively more when the earnings surprise is extreme. This result is consistent with event-based trading as stocks that either significantly beat earnings or spectacularly miss earnings can be characterized as those experiencing more extreme events, with a higher content of information. As I show later in Section 5.2, this event-based strategy of investing around earnings announcement days actually contributes to Robinhood investors' outperformance over the sample period.

4.2.3 Analyst Recommendation Revisions

I next corroborate Robinhood investors' preference for event-based trading by investigating their investment behavior around analyst recommendation revisions. I source the underlying data for analyst recommendations from the IBES Recommendations - Detail database. Recommendations are scored on a scale of 1 to 5, and include the categories {Strong Sell, Sell, Hold, Buy, Strong Buy}. For a given analyst-stock pair, I define a recommendation revision as an event day on which the given analyst changed their recommendation for the stock from their previous recommendation. For the purposes of this computation, the extent of the change is irrelevant: any move in the positive direction (an upgrade) is given a score of 1 and any move in the negative direction (a downgrade) is given a score of -1. For each stock-date pair, the aggregate analyst recommendation revision is then computed as the net number of upgrades less downgrades for the stock on that date across all analysts covering that stock. I merge the analyst recommendation revisions data with the main Robinhood dataset by linking IBES tickers with CRSP permnos according to the procedure outlined in Drechsler (2023). Specifically, the procedure primarily relies on CUSIP matching when available since it is more reliable. The remaining securities are classified via exchange ticker matching between IBES and CRSP, with additional date overlap and fuzzy company name matching layer checks.

Similarly to my analysis on earnings surprises, I perform a three-day short-window event study around days with analyst recommendation revisions. The analyst recommendation revisions are grouped into downgrades for all negative changes and upgrades for all positive changes. Although occasionally a rating upgrade or downgrade includes more than one level, the majority of rating revisions are one level upgrades or downgrades. Analyst recommendation revision days constitute a distinctly different subset of event days from the earnings announcement days, and recommendation revision days are well dispersed.

For the upgrade days, downgrade days, and control days with no rating revisions, I next compute the average contemporaneous three-day percent change in the number of Robinhood investors. As shown in Table 11, I find that both positive and negative rating changes elicit a positive investment response from Robinhood traders, consistent with event-based trading in response to analyst recommendation revisions.

4.3 The Influence of WallStreetBets

4.3.1 Popularity Analysis

I next investigate whether the popular investing subreddit WallStreetBets affects Robinhood investment. As detailed in Section 2, the WallStreetBets data comes from pushshift.io, covers the sample period May 2, 2018 - April 30, 2020, and includes data on post title, content, date posted, number of associated comments, the associated score, and number of awards received.

In order to test whether the WallStreetBets posts have an effect on Robinhood investment, I first construct four measures of popularity. The four measures are based on post mentions, comments, awards and scores respectively. I compute popularity for each date-ticker observation according to each of the popularity measures in the following manner:

1. First, I search through all of the posts and identify those that mention either stock tickers or company names. Both stock tickers and their associated company names are sourced from CRSP. Company names are further cleaned to remove words such as “CORP”, “HOLDINGS”, and “INC” since most posts use a short name for the company.[33] Stock tickers are typically mentioned in

[33] The full details of the company name matching algorithm between CRSP and WallStreetBets are provided in Appendix C.4.

capital letters in the posts, which is how I identify the mentions. Note that I remove observations for tickers which are common one letter, two letter, three letter or four letter words in the English language, in order to avoid counting them as tickers. At this point, I arrive at a dataset of posts, comments, awards and scores and the associated tickers mentioned in each post.

2. Next, I aggregate the data into date-ticker observations. In order to calculate the aggregate number of mentions of ticker i on date t, I sum the number of posts that contain that ticker or the associated company name on date t. For comments, awards, and scores I proceed slightly differently. To calculate the aggregate comments, awards, and scores for ticker i on date t, I sum the number of comments, awards and scores in all of the posts that mention ticker i on date t. After this aggregation, I am left with date-ticker observations which include the aggregate number of WallStreetBets mentions and associated comments, awards and scores. I merge this into my Robinhood dataset on ticker and date.

I then test whether popularity on the WallStreetBets platform drives Robinhood investment. In order to do this, I regress the percent change in the number of Robinhood investors on the four newly computed WallStreetBets stock popularity measures. As in the prior specifications, I also include controls, time fixed effects, and cluster the standard errors by security and time. Table 12 reports the results of the following regression:

$$\begin{aligned}\Delta \log NI_{i,t+1} = {} & \beta_1 \mathbb{1}_{i,t}^{PostMention} + \beta_2 ExcWinsComments_{i,t} + \beta_3 ExcWinsScore_{i,t} \\ & + \beta_4 Awards_{i,t} + \sum_{\tau=t-4}^{t} \beta_{5,\tau} \Delta \log NI_{i,\tau} + \gamma' Controls_{i,t} + \delta_{t+1} + \epsilon_{i,t+1}\end{aligned} \tag{14}$$

where $\mathbb{1}_{i,t}^{PostMention}$ is an indicator variable for whether stock i was mentioned on the WallStreetBets subreddit on date t; $ExcWinsComments$ is the aggregate number of comments for posts that mentioned stock i on date t, divided by the previous 10-day average of comments for the stock and winsorized at the 0.5% and 99.5% level to remove outliers; $ExcWinsScore_{i,t}$ is the previous day excess winsorized score of stock i; and $Awards_{i,t}$ is the previous day number of awards of stock i. The remaining variables were defined in prior specifications. Note also that while I take the excess winsorized value for the comment and score variables, I take the raw number of awards. The reason for this is that awards on the WallStreetBets platform are very infrequent, and there are no significant outliers.

As demonstrated in Table 12, each of the four WallStreetBets variables positively and statistically significantly predicts Robinhood investment the following day. This demonstrates that the peer network of investment advice sharing on WallStreetBets does indeed affect Robinhood investors' decision-making, and could be a driver of increased stock market participation. The popularity of a stock on the WallStreetBets subreddit positively predicts Robinhood investment into that stock.

When all four WallStreetBets variables are combined into a single regression as in Column (5) of Table 12, the strongest positive predictors of Robinhood investment are the post mention indicator variable and the number of excess winsorized comments, rather than the score or number of awards that a post received.

4.3.2 Sentiment Analysis

Interestingly, investors on the WallStreetBets forum appear to have rediscovered some age-old financial investing wisdom. One of these is represented by the diamond hands emoji in the WallStreetBets lexicon, which is indicative of avoiding panic and instead holding on to one's investments even in tough times. A common mistake investors make is to panic-sell during market crashes and subsequently miss out on the market recovery, which financial advisors the world over warn against. Robinhood investors, on the contrary, avoided the panic during the Covid-19 stock market crash of March 2020, and instead increased their holdings as demonstrated in Figure 8. This response during a large crisis and the widespread use of the diamond hands emoji points to the Robinhood investors having at least somewhat mastered this age-old financial wisdom of avoiding panic selling during market crashes. While this wisdom is most commonly used in the context of the entire market, for Robinhood investors it also extends to avoiding panic selling during rough times for individual stocks as well.

Along with the diamond hands emoji, WallStreetBets subreddit participants employ a host of other emojis, slang words, and abbreviations. Slang words include the popular word "HODL" used on the WallStreetBets forum to represent a similar concept of holding on to one's investments and avoiding selling or excessive trading. Initially a misspelling of the word "hold", investors have now popularized the slang term hodl in its place to mean to hold on to a stock long-term. Similarly to the diamond hands emoji, this term is in line with popular financial advice to avoid overtrading and keep investments for the long term.

Other common abbreviations used on the WallStreetBets platform include YOLO, which stands for the optimistic sentiment “you only live once”, and TLDR, which stands for “too long didn't read,” a shortcut to the main message of the post. Emojis used on the WallStreetBets platform also cover a wide range. Among the most prominent are the rocket ship emoji denoting the phrase “to the moon” and indicating optimism and the bear emoji representing short sellers or those with a negative market outlook. A sample of a dozen bullish and bearish posts are shown in Figure 14.

The standard Harvard sentiment dictionary commonly used in natural language processing (NLP) analysis does not cover this vernacular, and neither does the more financially relevant sentiment dictionary of Loughran and McDonald (2011) used for 10-K statement financial analysis. For this reason, I construct my own WallStreetBets sentiment dictionary using a simple, straightforward and systematic methodology.

The sentiment dictionary construction process works as follows. First, I identify 500 posts on the WallStreetBets subreddit that include a single ticker mention and either the word “buy” or “sell.” I next go through the 500 posts and manually verify that they do in fact convey bullish and bearish sentiment, respectively. If they do not, I drop the posts. Next, I separate the bullish sentiment posts from the bearish sentiment posts and use word frequency analysis to identify the most commonly occurring one-word, two-word, and three-word phrases among the bullish and bearish posts. Note that in this context emojis are treated as one-word phrases. I drop any words that are common connector words or do not have a strong positive or negative connotation. The resulting top 10 bullish phrases and top 10 bearish phrases used on the WallStreetBets platform are reported in Table 13.

Once I have the WallStreetBets dictionary, I use it to identify bullish and bearish posts in the following manner. For each post that includes a ticker mention, I search for the ten bullish and ten bearish phrases. If the post contains at least one bullish phrase and no bearish phrases, I classify it as a bullish sentiment post. If the post contains at least one bearish phrase and no bullish phrases, I classify it as a bearish sentiment post. Finally, if the post contains either no bullish and no bearish phrases or at least one of each, I classify it as a neutral sentiment post.

Once all the posts are classified, I aggregate sentiment to the ticker-date level by summing the sentiment scores of all posts that mention the ticker on that date. If this overall score is positive, I set the sentiment value for that ticker on that date equal to 1. If it is negative, I set it to -1. If it is 0, I leave the sentiment score as 0.

I use this measure of sentiment to test whether WallStreetBets sentiment can also help predict

Robinhood investment. Specifically, I employ the following regression:

$$\Delta \log NI_{i,t+1} = \beta_1 Sentiment_{i,t} + \sum_{\tau=t-4}^{t} \beta_{2,\tau} \Delta \log NI_{i,\tau} + \gamma' Controls_{i,t} + \delta_{t+1} + \epsilon_{i,t} \tag{15}$$

where $Sentiment_{i,t}$ is the sentiment of stock i at time t, and all the other variables in the regression retain their meaning from earlier specifications.

Table 14 reports the results, indicating that higher WallStreetBets sentiment for a stock predicts a positive and statistically significant increase in the number of Robinhood investors in that stock the following day. This demonstrates that not only the popularity of stocks on WallStreetBets but also the users' sentiment regarding those stocks helps explain Robinhood traders' investing decisions.

5 Portfolio Analysis

In this section, I investigate the performance of the representative Robinhood portfolio using the dollar method and share method of portfolio construction. I document significant outperformance of 14.8% to 21.4% annualized depending on the multifactor asset pricing model employed to measure alpha. I then decompose this outperformance by time period and strategy. I find that most of the outperformance occurred in 2020 and can be partially attributed to the buy-the-dip effect and event-based trading, but not to WallStreetBets. Finally, I analyze return predictability and show that Robinhood investment predicts market returns up to a one month horizon.

5.1 Performance

The focus of my paper lies in the study of the investing behaviors of the new class of retail investors, and the market design changes and behavioral biases that drive them. Nevertheless, it can also be instructive—particularly from a household finance and consumer welfare perspective—to investigate how much these drivers of behavior have helped or hurt the investors' performance. I find that over the sample period Robinhood investors significantly outperformed the market, and that this outperformance remains robust when controlling for multifactor models. Thus, performance constitutes yet another area in which the new class of retail investors has proven to be different from many people's preconceived beliefs and many earlier papers on retail investor underperformance.

In order to construct the aggregate representative Robinhood portfolio, I proceed via two methods.[34] The first is the dollar method, where I assume that each Robinhood investor represents an equal dollar amount investment in the stock. Without loss of generality, we can assume this investment amount is \$1. At each date, I compute the aggregate number of Robinhood investors in all stocks. I then compute the weight of the Robinhood aggregate representative portfolio in stock i at time t as follows:

$$w_{i,t}^{ptf,dollar} = \frac{NumInv_{i,t}}{\sum_j NumInv_{j,t}} \tag{16}$$

where $w_{i,t}^{ptf,dollar}$ is the dollar method Robinhood aggregate representative portfolio weight in stock i at time t, and $NumInv_{i,t}$ is the number of Robinhood investors in stock i at time t.

I also consider a second method of constructing the Robinhood aggregate representative portfolio. In this share method approach, each Robinhood investor in a stock represents a one share investment in that stock. The share method portfolio weights are calculated as:

$$w_{i,t}^{ptf,share} = \frac{NumInv_{i,t} * Prc_{i,t}}{\sum_j NumInv_{j,t} * Prc_{j,t}} \tag{17}$$

where $w_{i,t}^{ptf,share}$ is the share method Robinhood aggregate representative portfolio weight in stock i at time t, $NumInv_{i,t}$ is the number of Robinhood investors in stock i at time t, and $Prc_{i,t}$ is the price of stock i at time t.

In order to measure performance, I rely on daily stock level returns data and portfolio weights calculated according to these two methods. I assume a daily rebalancing frequency and compound the returns over time. Using both the dollar and share methods, I find that Robinhood investors actually outperformed the market and had a positive and statistically significant three factor model alpha, four factor model alpha, and six factor model alpha over the full sample period. The alpha was also economically large, at over 15% annualized for each of these three models.

Figure 15 depicts the performance of the Robinhood representative portfolio constructed using the dollar method, the performance of the Robinhood representative portfolio constructed using the share method, and the performance of the market as measured by the value-weighted return on all U.S. stocks. Both of the Robinhood representative portfolios considerably outperformed the market over the

[34] These methods are a natural way of aggregating Robinhood investors into a single portfolio, and are also employed by Welch (2022).

sample period. Note that the investment weights used each day for the representative portfolios were based on the dollar method and share method weights computed on the previous day, so there was no look-ahead bias and all information used was available at the time of investment. Another notable aspect that the performance chart reveals is the absence of significant drawdowns below the market level throughout the entire period.

Table 15 reports the performance breakdown of the dollar and share method Robinhood representative portfolios and the market portfolio by calendar year. From the table, it is clear that the majority of the outperformance can be attributed to the 2020 period, during which the Robinhood representative portfolios achieved a return more than four time larger than the corresponding market return. Performance of the Robinhood representative portfolio was roughly in line with the market in the 2018 period (-6.0% for the Robinhood portfolio using the dollar method and -6.3% for the Robinhood portfolio using the share method, versus -4.8% for the market) and modestly higher than the market in 2019 (38.3% for the Robinhood portfolio using the dollar method and 33.7% for the Robinhood portfolio using the share method, versus 30.5% for the market). From this annual breakdown, it is evident that Robinhood investors did not significantly underperform the market in any given year, and that most of their outperformance came during the 2020 period. A similar pattern emerges in Figure 16, which depicts the performance of the two Robinhood representative portfolios and the market portfolio for each year separately.

I also test the significance of the Robinhood representative portfolio's outperformance using the CAPM model, Fama-French three factor model (FF3), Fama-French-Carhart four factor model (FFC4), and Fama-French six factor model (FF6). The results for the Robinhood portfolio alpha and factor loadings using each of these asset pricing models are reported in Table 16. I find that the Robinhood portfolio alpha is positive and economically large in all four regression specifications, and significant at the 1% level for the three factor model, four factor model, and six factor model. In terms of economic magnitude, the estimated annualized alpha of the Robinhood portfolio is 14.8% using the CAPM, 21.4% using the three factor model, 15.9% using the four factor model, and 18.0% using the six factor model. These alphas are economically very large and statistically significant at the 1% level for three of the four models, which indicates that Robinhood investors may not only differ from previously studied investors with regards to their investing preferences but also with regards to their performance.

5.2 Decomposing Abnormal Returns

I next link the observed outperformance with the three key components underlying Robinhood investment that I identified earlier. My analysis reveals that both the buy-the-dip effect and event-based trading contributed significantly to the outperformance observed over the sample period, while a portfolio based on WallStreetBets did not yield statistically significant returns.

To test the performance of a strategy based on the buy-the-dip effect, I first denote a buy-the-dip stock as a (i) large-cap stock whose (ii) present day return falls in the lowest one-day return decile. In this way, each stock-date BTD event captures an extreme negative return of a large-cap company. I next compute the average future one-day return for BTD stocks and the average future one-day return for the control group of non-BTD stocks. The results are reported in Panel A of Table 17. Specifically, I find that the next-day returns of BTD stocks are on the order of 2.5x higher than the analogous mean return for the control group stocks.[35] This indicates that trading in response to the buy-the-dip effect could have contributed to Robinhood outperformance. However, one caveat is that the buy-the-dip strategy ceases to be profitable as soon as the holding period is extended to three days—in this case, BTD stocks perform significantly worse than the control group. The same observed patterns, namely the outperformance of the BTD strategy at a one-day horizon and underperformance using a three-day holding period, also hold out of sample over the longer May 2015 - June 2023 period. This time period includes data from three years prior to the Robinhood sample period start date and nearly three years after the Robinhood sample period end date.

To determine the performance of an event-based strategy over the sample period, I focus on earnings announcement days. Panel B of Table 17 reports the results. Three-day returns around earnings announcement days generally follow the pattern predicted, namely that companies that miss earnings face negative three-day event returns while companies that beat earnings are rewarded with positive three-day returns. However, a more relevant and less immediately obvious finding for the event-based strategy is that three-day returns around days when companies release earnings—aggregated and averaged across all types of positive and negative earnings surprises—yield significantly higher three-day returns than the control group when no earnings are announced. For instance, over the Robinhood sample period the average three-day return for stocks with earnings announcements was 51.12bps while

[35] This is computed as $0.1060\%/0.0422\% \approx 2.5$ from data in Table 17 for BTD and Non-BTD one-day mean forward returns over the Robinhood sample period.

that of stocks that did not announce earnings was 14.40bps. The same pattern holds when focusing only on contemporaneous one-day returns, as well as out of sample for the entire period of May 2015 - June 2023. These robust findings underscore the equity risk premium type reward associated with earnings announcement days. This is in line with the event-based trading driver of Robinhood investment, and likely contributes to their outperformance. This finding also underscores the possibility that the new class of retail investors may be benefiting from the risk premium associated with announcement event days (e.g., Savor and Wilson (2014)) and in particular the earnings announcer premium (Savor and Wilson (2016)), even as their investment behaviors may be influenced by behavioral effects such as bounded rationality, limited attention, and the law of small numbers.

Finally, I also construct a portfolio based on WallStreetBets popularity mentions where the weight of a stock in the portfolio is equal to the stock's popularity divided by the aggregate popularity of all stocks being discussed on the platform. However, I do not observe outperformance for the WallStreetBets strategy and conclude that it was not a significant driver of outperformance over the sample period.

5.3 Return Predictability

Finally, I investigate the relationship between Robinhood investment and future returns. I analyze the predictability of weekly returns, and document predictability up to a one month horizon. Specifically, my panel regression results suggest that Robinhood investment is related to future returns in a statistically significant manner, and that the predictability remains intact even after controlling for past daily, weekly, and monthly returns, momentum, and other control variables. The full regression results of this analysis are reported in Table 18, and suggest that Robinhood investment adds to the prediction of returns provided by well-known factors such as size and the book-to-market ratio. Insofar as this return predictability continues, this is supporting evidence of Robinhood investment having the capacity to improve market efficiency and price discovery up to a one month horizon.

6 Policy Implications

In this section, I discuss several policy implications associated with the recent rise in Robinhood investment. Specifically, I focus on the implications for stock market participation and individual investor safeguards. I illustrate how my findings lend support to the participation costs channel and peer effects channel in explaining the stock market participation puzzle, and do not offer evidence for the preference-based channel and risk-based channel. I discuss the impact of recent market design changes on individual investors and highlight remaining impediments including asset class limitations and accredited investor status.

6.1 Stock Market Participation

One longstanding puzzle in the finance literature is the stock market participation puzzle, namely that a large fraction of the population in the U.S. and worldwide does not hold any equity investments (Mankiw and Zeldes (1991), Haliassos and Bertaut (1995)). This finding is true both for the U.S. and worldwide, with the participation rate in the U.S. just below 50% and even lower participation rates for other countries globally (Guiso, Sapienza, and Zingales (2008), Christelis, Georgarakos, and Haliassos (2013)). When considering the fraction of households that invest in the stock market directly—i.e., not including those who hold equity investment indirectly through investment vehicles such as retirement plans—the proportion of individuals holding stocks decreases further to the order of 20% (Badarinza, Campbell, and Ramadorai (2016)).

To see why the low stock market participation rate is a puzzle, consider the following simple model. Consider a static portfolio choice problem with a single period and two assets, a riskless asset with return R_f and a risky asset with return $R_f + \tilde{x}$. Let the individual be endowed with wealth W_0.

After one period, the individual receives $W_1 = \theta(1 + R_f + \tilde{x}) + (W_0 - \theta)(1 + R_f) = W_0(1 + R_f) + \theta\tilde{x}$, where θ is the dollar amount of wealth invested in the risky asset. The investor wishes to maximize expected utility, and the optimization problem is:

$$\max_{\theta} V(\theta) = \max_{\theta} \mathrm{E}[u(W_0(1 + R_f) + \theta\tilde{x})] \tag{18}$$

The first-order condition is then $\mathrm{E}[u'(W_0(1 + R_f) + \theta\tilde{x})\tilde{x}] = 0$. Evaluating the first-order condition at zero investment in the risky asset ($\theta = 0$) yields $u'(W_0(1 + R_f))\mathrm{E}[\tilde{x}]$. Note that for any increasing utility function, the first portion of the expression, $u'(.)$, is positive. Therefore, the sign of the expression must be the same as the sign of $\mathrm{E}[\tilde{x}]$. As long as the expected risk premium for investing in the riksy asset is positive ($\mathrm{E}[\tilde{x}] > 0$), the derivative is positive and implies that a positive amount of wealth should be invested in the risky asset.

Several possible explanations have been proposed in response to the stock market participation puzzle. These explanations have tended to fall into four main categories: utility function explanations, participation costs, risk-based explanations, and peer effects. The preference-based explanations have focused on household preferences that depart from expected utility. The participation costs explanations have included both pecuniary and informational costs, as well as fixed costs and ongoing participation costs. Risk-based explanations have investigated the labor income risk faced by households, and the extent to which labor income is correlated with stock market returns. Finally, peer effects explanations have focused on the influence of peers in driving household investment decisions.

In the context of Robinhood and the rise of the new class of retail traders, increased stock market participation has been accompanied by four major changes. The first is the advent of zero commission trading. The second is the relaxation of capital constraints through both zero account minimums and the ability to trade fractional shares. The third is the increased free time and decreased outside option value associated with lockdowns, and the fourth is the capital injection through the stimulus checks program. The first two represent thus far permanent shifts in the market design landscape, while the last two are temporary. Nevertheless, while lockdown directives and stimulus check disbursements are not ongoing changes, they have likely had a lasting effect on retail stock market participation rates, particularly due to demonstrated inertia in individual decision-making.

Taken together, these four changes are most consistent with the participation cost and peer effects channels. The elimination of commissions trading costs, relaxation of investment minimums, introduction of fractional share trading, and additional liquidity through stimulus checks have all directly lowered participation costs. Meanwhile, increased free time to learn from peers during lockdowns and my findings on the influence of WallStreetBets popularity and sentiment for Robinhood investment point to the importance of the peer effects channel. Consequently, the recent rise in retail participation can most closely be linked to these two types of explanations. On the other hand, preference-based explanations or risk-based explanations are less likely to explain the recent rise in stock market participation rates. It is harder to argue that the four changes have impacted the form of investors' utility functions in a way that more closely mirrors expected utility theory or that the correlation of individuals' labor income with stock returns has dramatically decreased. Within the participation cost and peer effects channels, future work can focus on quantifying the exact impact of each channel on stock market participation rates. In particular, studies could attempt to estimate the participation cost elasticity of demand for stock market participation or the quantity of stock market exposure demanded conditional on participation for the new class of retail investors.

From a policy perspective, the above discussion supports the view that policies which seek to lower stock market participation costs or increase the peer influence of stock market participants would be effective instruments for further increasing stock market participation among U.S. households. It is also important to keep in mind that household-level differences in expected returns are influenced by past experiences, i.e. stock market returns experienced by investors over their lives (Malmendier and Nagel (2011)). This indicates that the new class of retail investors, many of whom entered the stock market in 2018-2020 during a period of relatively high returns, are likely to have formed relatively optimistic beliefs with regards to future stock market outlook. In turn, this can help influence these investors to remain invested in the stock market and mark a persistent shift in the U.S. stock market participation rate. Policies focused on increasing household participation in the financial markets can attempt to focus their efforts more on stable periods, since encouraging investment during a turbulent or recessionary period can have long-lasting counterproductive effects.

6.2 Individual Investor Safeguards

As discussed, the retail investing landscape has undergone a dramatic shift in part due to market design changes that have lowered participation costs and increased investing accessibility for retail traders. In turn, this has successfully increased stock market participation. According to the Survey of Consumer Finance (SCF), direct equity ownership increased by 38.2% over the past three years, rising from 15.2% in 2019 to 21.0% in 2022.[36] A natural question then arises: is this increased participation in the equity markets beneficial for these individual investors, or should more safeguards be put in place? What does this shock to stock market participation tell us about other asset classes and limitations placed on individual investors?

My results on the performance of Robinhood individual investors point to a promising picture of individual investors' aggregate ability to navigate the financial markets. Over May 2018 - August 2020, the representative Robinhood portfolio outperformed the market, with the outperformance remaining robust after controlling for multifactor models. Over the sample period investors achieved an alpha of

[36] Source: Survey of Consumer Finances (SCF).

over 15% annualized. If the sample is restricted to exclude the exceptional Covid pandemic period, then the Robinhood portfolio still outperforms the market. Furthermore, I find that Robinhood investment predicts future market returns up to a one month horizon over the sample period. Insofar as that continues, it supports the view that on average, Robinhood investors can help improve market efficiency and the price discovery process.[37]

The past decade has also witnessed increased accessibility for individual investors to trade in additional asset classes. For example, Robinhood allowed investors to trade options in 2017 and rolled out cryptocurrency trading throughout 2019. As related literature shows (e.g. Kogan et al. (2023)), these new asset classes have seen significant uptake by retail investors. My research supports the view that allowing for easier trading in these asset classes has allowed retail investors to more easily express any lottery preferences there, and resulted in lower lottery stock investing in direct equity markets. Recent research shows less than stellar performance by retail traders in the options market (e.g. Beckmeyer et al. (2023)), but also notes that much of the underperformance comes from exorbitant bid-ask spreads—an aspect of options trading that is likely not as frequently faced by institutional investors who have the option to call up personal brokers to negotiate OTC trades and do not need to rely on online brokerages for options trading.

Where does the way forward lead? While retail investors have gained access to low-cost trading in the equity markets, effective trading costs on derivative instruments remain high and other oftentimes lucrative asset classes such as venture capital, private equity, and hedge fund investments remain altogether out of reach. As noted by the famed Vanguard Group founder and index mutual fund pioneer John C. Bogle, an optimal portfolio relies on asset allocations to a diverse set of investments. For instance, Jack R. Meyer, former president of Harvard Management Company—who tripled the Harvard University endowment fund from $8bn to $27bn—advises investors "First, get diversified. Come up with a portfolio that covers a lot of asset classes." Following this advice, in 2010 the Harvard Management Company (HMC) allocated a hefty 60% policy weight to exotic asset classes such as private equity and real estate, more than domestic and international stocks and bonds combined. Yet exotic asset classes, and their returns, remain largely out of reach for individual investors.

[37]One important caveat to note are the special episodes of coordination among retail investors, particularly those around so-called "meme stocks." Barber et al. (2022) show that the very top herding stocks of Robinhood investors underperform, and Allen et al. (2023) show that 13 meme stock episodes based on social media coordination impeded market quality. Therefore, these rarer special episodes of large-scale market coordination may warrant special policy scrutiny.

Going forward, limitations on retail investor participation in the form of accredited investor status—a status that relies on net worth and income levels rather than on financial educational attainment in the U.S.—may be reassessed to allow for more equal and inclusive access to participation in the financial markets. Increasing the access of lower income households to these investments, with the proper financial literacy in place, also has the potential to help lower inequality.

7 Conclusion

The Robinhood platform was founded in 2013 as a pioneer of zero-commission trading, and has since revolutionized retail investing. Attracting over 18 million funded accounts with roughly half being first-time investors, Robinhood has appeared to make headway towards making investing easier, cheaper, and more accessible for millions of users who were heretofore absent from the stock market. Indeed, in facilitating retail trading Robinhood has shifted the market equilibrium among all major U.S. online brokerages and heralded a new age for a new class of relatively young, small, and inexperienced retail investors.

In this paper, I determine the differences in investing behavior of the new class of retail investors relative to previously studied individual investors. I find that in contrast to prior literature, Robinhood investors do not exhibit a preference for investing in lottery stocks, small stocks, or value stocks. I explain the absence of investment in lottery stocks by employing a regression discontinuity in time design around the date of the introduction of fractional share trading on December 12, 2019. Meanwhile, I also show that the differences in value stock and small-cap stock investment are consistent with extrapolative beliefs in style investing.

I also contribute by documenting three main components that help explain Robinhood investment. The first is a novel buy-the-dip effect, where Robinhood investors exhibit a preference for investing in large stocks that have experienced a negative extreme return, or a dip. I hypothesize that this is due to ex-ante brand affinity for more well-known and typically larger companies, such that Robinhood investors wait for what they consider as the right moment before investing. The buy-the-dip effect is not present for small-cap stocks. Along with evidence based on quintile analysis and regression analysis, several of the top Robinhood holdings also illustrate the buy-the-dip effect. I propose a model based on the law of small numbers bias to explain these findings.

The second component that drives Robinhood investment is event-based trading. In particular, Robinhood investors invest more following earnings announcements, large earnings surprises, analyst recommendation revisions, and extreme returns, volatility, and volume traded. This is consistent with bounded rationality and limited attention theories, and I rule out the private information channel.

Finally, I also provide evidence that stocks with higher popularity and more bullish sentiment on the WallStreetBets platform exhibit greater trading by Robinhood investors. This indicates that the degree to which these investors are well-connected and share investing tips amongst themselves plays a role in their investment choices. To supplement this analysis, I construct a novel financial forum dictionary based on the WallStreetBets platform vernacular incorporating slang terms and emoji data. This dictionary can be used in future research on analogous financial social network platforms.

References

Allen, F., M. Haas, E. Nowak, M. Pirovano, and A. Tengulov (2023): "Squeezing Shorts Through Social Media Platforms," *Working Paper*.

Amaya, D., P. Christoffersen, K. Jacobs, and A. Vasquez (2015): "Does Realized Skewness Predict the Cross-Section of Equity Returns?" *Journal of Financial Economics*, 118, 135–167.

Amihud, Y. (2002): "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects," *Journal of Financial Markets*, 5, 31–56.

Ardia, D., C. Aymard, and T. Cenesizoglu (2023): "Fast and Furious: An Intraday Analysis of Robinhood Users' Trading Behavior," *Working Paper*.

Badarinza, C., J. Y. Campbell, and T. Ramadorai (2016): "International Comparative Household Finance," *Annual Review of Economics*, 8, 111–144.

Badrinath, S. G. and W. G. Lewellen (1991): "Evidence on Tax-Motivated Securities Trading Behavior," *Journal of Finance*, 46, 369–382.

Bali, T. G., N. Cakici, and R. F. Whitelaw (2011): "Maxing Out: Stocks as Lotteries and the Cross-Section of Expected Returns," *Journal of Financial Economics*, 99, 427–446.

Barber, B. M., X. Huang, T. Odean, and C. Schwarz (2022): "Attention Induced Trading and Returns: Evidence from Robinhood Users," *Journal of Finance*, 77, 3141–3190.

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

——— (2001): "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment," *Quarterly Journal of Economics*, 116, 261–292.

——— (2008): "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors," *Review of Financial Studies*, 21, 785–818.

——— (2013): "The Behavior of Individual Investors," *Handbook of the Economics of Finance*, 2, 1533–1570.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2015): "X-CAPM: An Extrapolative Capital Asset Pricing Model," *Journal of Financial Economics*, 115, 1–24.

——— (2018): "Extrapolation and Bubbles," *Journal of Financial Economics*, 129, 203–227.

Barberis, N. and M. Huang (2008): "Stocks as Lotteries: The Implications of Probability Weighting for Security Prices," *American Economic Review*, 98, 2066–2100.

Barberis, N. and A. Shleifer (2003): "Style Investing," *Journal of Financial Economics*, 68, 161–199.

Barrot, J.-N., R. Kaniel, and D. Sraer (2016): "Are Retail Traders Compensated for Providing Liquidity?" *Journal of Financial Economics*, 120, 146–168.

Bastianello, F. and P. Fontanier (2022): "Expectations and Learning from Prices," *Working Paper*.

Beckmeyer, H., N. Branger, and L. Gayda (2023): "Retail Traders Love 0DTE Options... But Should They?" *Working Paper*.

Bernile, G., V. Bhagwat, and P. R. Rau (2017): "What Doesn't Kill You Will Only Make You More Risk-Loving: Early-Life Disasters and CEO Behavior," *Journal of Finance*, 72, 167–206.

Boehmer, E., C. M. Jones, X. Zhang, and X. Zhang (2021): "Tracking Retail Investor Activity," *Journal of Finance*, 76, 2249–2305.

Carhart, M. M. (1997): "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52, 57–82.

Christelis, D., D. Georgarakos, and M. Haliassos (2013): "Differences in Portfolios Across Countries: Economic Environment Versus Household Characteristics," *Review of Economics and Statistics*, 95, 220–236.

Cookson, A., R. Lu, W. Mullins, and M. Niessner (2022): "The Social Signal," *Working Paper*.

Cookson, J. A., J. E. Engelberg, and W. Mullins (2023): "Echo Chambers," *Review of Financial Studies*, 36, 450–500.

Cooper, M. J., H. Gulen, and M. J. Schill (2008): "Asset Growth and the Cross-Section of Stock Returns," *Journal of Finance*, 63, 1609–1651.

Cutler, D. M., J. M. Poterba, and L. H. Summers (1990): "Speculative Dynamics and the Role of Feedback Traders," *American Economic Review*, 80, 63–68.

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

Da, Z., V. W. Fang, and W. Lin (2023): "Fractional Trading," *Working Paper*.

Davis, J. L., E. F. Fama, and K. R. French (2000): "Characteristics, Covariances, and Average Returns: 1929 to 1997," *Journal of Finance*, 55, 389–406.

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990): "Positive Feedback Investment Strategies and Destabilizing Rational Speculation," *Journal of Finance*, 45, 379–395.

Drechsler, Q. F. S. (2023): "Python Programs for Empirical Finance," *https://www.fredasongdrechsler.com*.

Eaton, G. W., T. C. Green, B. Roseman, and Y. Wu (2022): "Retail Trader Sophistication and Stock Market Quality: Evidence from Brokerage Outages," *Journal of Financial Economics*, 146, 502–528.

Fama, E. F. and K. R. French (1992): "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47, 427–465.

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

Fedyk, A. (2022): "Front Page News: The Effect of News Positioning on Financial Markets," *Journal of Finance (forthcoming)*.

Field, L. C. and J. M. Karpoff (2002): "Takeover Defenses of IPO Firms," *Journal of Finance*, 57, 1857–1889.

Foucault, T., D. Sraer, and D. J. Thesmar (2011): "Individual Investors and Volatility," *Journal of Finance*, 66, 1369–1406.

Garcia, D., X. Hu, and M. Rohrer (2023): "The Colour of Finance Words," *Journal of Financial Economics*, 147, 525–549.

Garrett, T. and R. Sobel (1999): "Gamblers Favor Skewness, Not Risk: Further Evidence From United States' Lottery Games," *Economic Letters*, 63, 85–90.

Greenwood, R., T. Laarits, and J. Wurgler (2023): "Stock Market Stimulus," *Review of Financial Studies*, 36, 4082–4112.

Greenwood, R. and A. Shleifer (2014): "Expectations of Returns and Expected Returns," *Review of Financial Studies*, 27, 714–746.

Grinblatt, M. and M. Keloharju (2000): "The Investment Behavior and Performance of Various Investor Types: A Study of Finland's Unique Data Set," *Journal of Financial Economics*, 55, 43–67.

——— (2001): "What Makes Investors Trade?" *Journal of Finance*, 56, 589–616.

Guiso, L., P. Sapienza, and L. Zingales (2008): "Trusting the Stock Market," *Journal of Finance*, 63, 2557–2600.

Haliassos, M. and C. C. Bertaut (1995): "Why Do So Few Hold Stocks?" *Economic Journal*, 105, 1110–1129.

Harvey, C. R. and A. Siddique (2000): "Conditional Skewness in Asset Pricing Tests," *Journal of Finance*, 55, 1263–1295.

Hu, D., C. M. Jones, V. Zhang, and X. Zhang (2023): "The Rise of Reddit: How Social Media Affects Retail Investors and Short-Sellers' Roles in Price Discovery," *Working Paper*.

Jegadeesh, N. and S. Titman (1993): "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance*, 48, 65–91.

Kahneman, D. and A. Tversky (1971): "Belief in the Law of Small Numbers," *Psychological Bulletin*, 76, 105–110.

Kalda, A., B. Loos, A. Previtero, and A. Hackethal (2021): "Smart(Phone) Investing? A within Investor-Time Analysis of New Technologies and Trading Behavior," *Working Paper*.

Kaniel, R., S. Liu, G. Saar, and S. Titman (2012): "Individual Investor Trading and Return Patterns around Earnings Announcements," *Journal of Finance*, 67, 639–680.

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

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

Kogan, S., I. Makarov, M. Niessner, and A. Schoar (2023): "Are Cryptos Different? Evidence from Retail Trading," *NBER*.

Kraus, A. and R. H. Litzenberger (1976): "Skewness Preference and the Valuation of Risk Assets," *Journal of Finance*, 31, 1085–1100.

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

Laarits, T. and M. Sammon (2023): "The Retail Habitat," *Working Paper*.

Liao, J., C. Peng, and N. Zhu (2022): "Extrapolative Bubbles and Trading Volume," *Review of Financial Studies*, 35, 1682–1722.

Loughran, T. and B. McDonald (2011): "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *Journal of Finance*, 66, 35–65.

Loughran, T. and J. Ritter (2004): "Why Has IPO Underpricing Changed over Time?" *Financial Management*, 33, 5–37.

Luo, C., E. Ravina, M. Sammon, and L. M. Viceira (2022): "Retail Investors' Contrarian Behavior Around News, Attention, and the Momentum Effect," *Working Paper*.

Malmendier, U. and S. Nagel (2011): "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?" *Quarterly Journal of Economics*, 126, 373–416.

Malmendier, U., G. Tate, and J. Yan (2011): "Overconfidence and Early-Life Experiences: The Effect of Managerial Traits on Corporate Financial Policies," *Journal of Finance*, 66, 1687–1733.

Mankiw, N. G. and S. P. Zeldes (1991): "The Consumption of Stockholders and Nonstockholders," *Journal of Financial Economics*, 29, 97–112.

Markowitz, H. (1952a): "Portfolio Selection," *Journal of Finance*, 7, 77–91.

——— (1952b): "The Utility of Wealth," *Journal of Political Economy*, 60, 151–158.

Moss, A., J. P. Naughton, and C. Wang (2023): "The Irrelevance of ESG Disclosure to Retail Investors: Evidence from Robinhood," *Management Science*, 70, 2023–2704.

Novy-Marx, R. (2013): "The Other Side of Value: The Gross Profitability Premium," *Journal of Financial Economics*, 108, 1–28.

Odean, T. (1998): "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance*, 53, 1775–1798.

——— (1999): "Do Investors Trade Too Much?" *American Economic Review*, 89, 1279–1298.

Ozik, G., R. Sadka, and S. Shen (2021): "Flattening the Illiquidity Curve: Retail Trading During the COVID-19 Lockdown," *Journal of Financial and Quantitative Analysis*, 56, 2356–2388.

Pagano, M. S., J. Sedunov, and R. Velthuis (2021): "How Did Retail Investors Respond to the COVID-19 Pandemic? The Effect of Robinhood Brokerage Customers on Market Quality," *Finance Research Letters*, 43, 1–11.

Pedersen, L. H. (2022): "Game On: Social Networks and Markets," *Journal of Financial Economics*, 146, 1097–1119.

Polkovnichenko, V. (2005): "Household Portfolio Diversification: A Case for Rank-Dependent Preferences," *Review of Financial Studies*, 18, 1467–1502.

Rabin, M. (2002): "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics*, 117, 775–816.

Rabin, M. and D. Vayanos (2010): "The Gambler's and Hot-Hand Fallacies: Theory and Applications," *Review of Economic Studies*, 77, 730–778.

Savor, P. and M. Wilson (2014): "Asset Pricing: A Tale of Two Days," *Journal of Financial Economics*, 113, 171–201.

——— (2016): "Earnings Announcements and Systematic Risk," *Journal of Finance*, 71, 83–138.

Seasholes, M. S. and G. Wu (2007): "Predictable Behavior, Profits, and Attention," *Journal of Empirical Finance*, 14, 590–610.

Stein, R. (2020): "The Top 5 Predictable Effects of New Entries in Robinhood's '100 Most Popular' List," *Working Paper*.

Tetlock, P. C. (2007): "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *Journal of Finance*, 62, 1139–1168.

Tversky, A. and D. Kahneman (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

Uettwiller, A. (2022): "Retail Investor Heterogeneity: Evidence from WallStreetBets," *Working Paper*.

van der Beck, P. and C. Jaunin (2023): "The Equity Market Implications of the Retail Investment Boom," *Working Paper*.

Welch, I. (2022): "The Wisdom of the Robinhood Crowd," *Journal of Finance*, 77, 1489–1527.

Tables

Table 1
Top 10 U.S. Stock Holdings of Robinhood Investors

This table shows the top 10 U.S. stocks by the number of Robinhood investors as of August 13, 2020, the Robintrack sample end date. Companies highlighted in blue include those that experienced a drawdown of at least 50% over the course of the Robintrack sample period, May 2, 2018 - August 13, 2020. The high representation of such companies in the top holdings list illustrates the buy-the-dip effect.

Company Name	Ticker	Number of Investors	Maximum Drawdown
Ford Motor	F	912,378	-62.94 %
General Electric	GE	857,478	-61.82 %
Apple	AAPL	726,024	-38.52%
Microsoft	MSFT	653,687	-28.04%
American Airlines	AAL	638,246	-79.44 %
Disney	DIS	598,684	-43.11%
Delta Air Lines	DAL	572,649	-69.18 %
Tesla	TSLA	563,974	-60.63 %
Gopro	GPRO	475,045	-73.38 %
Amazon	AMZN	427,221	-34.10%

Table 2
Representative Portfolio Sector Weights

This table shows the Standard Industry Classification (SIC) division weights (Panel A) and the North American Industry Classification System (NAICS) sector weights (Panel B) for the market portfolio, Robinhood (dollar method) portfolio, and Robinhood (share method) portfolio as of August 13, 2020 (the Robintrack sample period end date). Historical SIC information is from CRSP. SIC Division classification codes are from https://www.osha.gov/data/sic-manual and NAICS Sector classification codes are from the U.S. Census.

Panel A - SIC

SIC Division	Market Portfolio Weight	Robinhood (Dollar Method) Portfolio Weight	Robinhood (Share Method) Portfolio Weight
Agriculture, Forestry, and Fishing	0.08%	0.02%	0.00%
Construction	0.45%	0.16%	0.13%
Finance, Insurance, and Real Estate	12.89%	5.42%	1.90%
Manufacturing	38.75%	37.04%	22.81%
Mining	1.95%	6.61%	0.54%
Public Administration	0.00%	0.05%	0.00%
Retail Trade	6.77%	7.66%	4.04%
Services	29.00%	26.85%	66.96%
Transportation, Comm.s, Electric, Gas, & Sanitary	9.05%	15.31%	3.40%
Wholesale Trade	1.06%	0.88%	0.22%

Panel B - NAICS

NAICS Sector	Market Portfolio Weight	Robinhood (Dollar Method) Portfolio Weight	Robinhood (Share Method) Portfolio Weight
Accommodation and Food Services	1.64%	3.10%	1.39%
Administrative and Support Services	1.11%	0.52%	0.20%
Agriculture, Forestry, Fishing and Hunting	0.07%	0.01%	0.00%
Arts, Entertainment, and Recreation	0.23%	0.87%	0.25%
Construction	0.49%	0.16%	0.10%
Educational Services	0.05%	0.06%	0.01%
Finance and Insurance	12.60%	4.76%	2.02%
Health Care and Social Assistance	0.78%	0.63%	0.26%
Information	22.79%	19.38%	16.82%
Manufacturing	40.35%	46.08%	40.57%
Mining, Quarrying, and Oil and Gas Extraction	1.26%	4.87%	0.26%
Other Services	0.14%	0.23%	0.03%
Professional, Scientific, and Technical Services	1.26%	0.57%	0.13%
Real Estate and Rental and Leasing	0.99%	1.83%	2.72%
Retail Trade	10.08%	5.03%	33.13%
Transportation and Warehousing	2.13%	10.23%	1.74%
Utilities	3.02%	0.78%	0.26%
Wholesale Trade	1.00%	0.88%	0.10%

Table 3
Lottery Stock, Non-Lottery Stock, and Other Stock Characteristics

This table reports the mean daily characteristics of lottery stocks, non-lottery stocks, and other stocks. Characteristics are measured and averaged over the entire May 2, 2018 - August 13, 2020 sample period. On each day, a lottery stock is defined as a stock that falls into the lowest 50th price percentile, the highest 50th idiosyncratic volatility percentile, and the highest 50th idiosyncratic skewness percentile. A non-lottery stock is defined as a stock that falls into the highest 50th price percentile, the lowest 50th idiosyncratic volatility percentile, and the lowest 50th idiosyncratic skewness percentile. Other stocks include all remaining stocks that do not fall into the lottery stock and non-lottery stock definition categories. Market beta, SMB beta, HML beta, and MOM beta are computed as the loadings on the Fama-French-Carhart four factor model computed using daily returns data over the past year. The book-to-market ratio and past 12-month return less previous month return variables are winsorized at the [0.5%, 99.5%] level.

Measure	Lottery Stocks	Non-Lottery Stocks	Other Stocks
Number of stocks	728	760	2,148
Total volatility (ann.)	92.08%	32.80%	53.17%
Idiosyncratic volatility (ann.)	88.60%	24.43%	44.29%
Total skewness	1.56	-0.56	0.07
Idiosyncratic skewness	1.72	-0.65	0.27
Stock price	5.71	331.42	96.90
Market beta	0.83	0.91	0.89
Firm size (mm)	402	20,627	6,100
SMB beta	0.85	0.47	0.77
Book-to-market ratio	0.82	0.47	0.59
HML beta	-0.09	0.23	0.06
Past 12-month less previous month return	-11.21%	5.76%	1.40%
MOM beta	-0.26	0.06	-0.04
Amihud illiquidity	0.89	0.01	0.22
Annual volume turnover (shares, mm)	0.99	1.53	1.15
Firm age (years)	22.87	43.59	25.87
Percentage dividend paying	12.06%	72.41%	40.77%

Table 4

Lottery Stock, Value, and Size Investment Preferences

This table reports the panel regression estimates with time fixed effects in columns (1), (2), (3), and (4) and the Fama and MacBeth (1973) cross-sectional regression estimates in columns (5) and (6) for the regressions of the number of Robinhood investors on lottery stock characteristics, size, and value. The dependent variable is computed as the log winsorized number of Robinhood investors in stock i at time t. All independent variables are measured at the end of the previous day. The control variables include momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, dividends per share, an indicator variable for a dividend-paying firm, firm age, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. The independent variables of interest are standardized. Regressions include time fixed effects and standard errors are clustered by time and security.

	Number of Robinhood Investors					
	(1)	(2)	(3)	(4)	(5)	(6)
Stock price	4.60			8.30^{***}	4.65^{***}	7.86^{***}
	(3.01)			(1.81)	(0.16)	(0.07)
Idiosyncratic volatility	0.54^{***}			0.22^{***}	0.56^{***}	0.25^{***}
	(0.05)			(0.04)	(0.01)	(0.00)
Idiosyncratic skewness	-0.13^{***}			-0.05^{***}	-0.12^{***}	-0.04^{***}
	(0.02)			(0.01)	(0.00)	(0.00)
ln(Size)		0.53^{***}		0.12^{*}		0.10^{***}
		(0.03)		(0.07)		(0.00)
Book-to-market ratio			-0.26^{***}	-0.14^{***}		-0.14^{***}
			(0.03)	(0.02)		(0.00)
Controls	No	No	No	Yes	No	Yes
R-squared	0.11	0.12	0.07	0.67	0.07	0.67
Time fixed effects	Yes	Yes	Yes	Yes	-	-
Number of observations	1,655,058	1,823,564	1,702,386	1,565,496	1,655,058	1,565,496

Table 5
Extrapolative Beliefs and Demand for Value and Size

This table reports the one year, three year, and five year cumulative value (HML) and size (SMB) factor returns in the years leading up to the Barber and Odean (2000) sample (Panel A) and the Robinhood sample (Panel B). The significant difference in returns points to extrapolative beliefs in the allocation of capital to style investment types among retail investors.

Panel A - Large U.S. Brokerage House

	Value			**Size**		
	HML_1yr_ret	HML_3yr_ret	HML_5yr_ret	SMB_1yr_ret	SMB_3yr_ret	SMB_5yr_ret
1991	-10.00	-2.45	1.36	-15.20	-21.20	-39.41
1992	-11.53	-23.09	-13.99	10.64	-15.89	-24.73
1993	22.57	-0.46	5.56	5.51	0.27	-6.75
1994	15.20	24.73	8.15	5.36	25.48	-4.63
1995	-0.73	40.68	14.05	-3.16	7.30	1.82
1996	3.33	18.18	27.95	-5.90	-3.99	14.34
Mean	3.14	9.60	7.18	-0.46	-1.34	-9.89

Panel B - Robinhood

	Value			**Size**		
	HML_1yr_ret	HML_3yr_ret	HML_5yr_ret	SMB_1yr_ret	SMB_3yr_ret	SMB_5yr_ret
2018	-11.03	-3.75	-3.69	-4.45	-2.68	-4.25
2019	-9.32	-2.59	-14.09	-5.04	-2.28	-14.37
2020	-10.02	-27.09	-21.60	-6.37	-15.34	-13.32
Mean	-10.13	-11.14	-13.12	-5.28	-6.76	-10.64
Diff.	13.27	20.74	20.30	4.83	5.43	0.75
t-stat	1.63	1.42	2.26	0.85	0.52	0.06

Table 6

3-Day Change in Number of Robinhood Investors by Cumulative Abnormal Return (CAR) Quintile

This table shows the average three-day percent change in the number of Robinhood investors for stocks in a given cumulative abnormal return (CAR) quintile. CAR is computed by compounding the contemporaneous three-day excess return net of the risk-free rate. Quintile 1 represents the lowest CAR quintile and Quintile 5 represents the highest CAR quintile. Small Cap includes all observations for stocks with a market cap less than or equal to $2bn, Mid Cap includes all observations for stocks with a market cap between $2bn and $10bn, and Large Cap includes all observations for stocks with a market cap over $10bn. The Overall column includes observations across all market caps. The Observation Count column reports the number of observations used to compute the quintile averages in the Overall column.

CAR Quintile	Small Cap	Mid Cap	Large Cap	Overall	Observation Count
Quintile 1	1.6%	2.0%	2.4%	1.8%	401,674
Quintile 2	0.4%	0.3%	0.6%	0.4%	338,407
Quintile 3	0.5%	0.3%	0.4%	0.4%	314,635
Quintile 4	0.5%	0.3%	0.3%	0.4%	338,582
Quintile 5	3.1%	1.5%	1.1%	2.4%	400,938

Table 7
Buy-the-Dip Effect, Small Cap vs. Large Cap

This table reports the results of two regressions of the percent change in the number of Robinhood investors on (i) Q1 or Q5 Return ($\mathbb{1}_{i,t}^{Q1or5}$), an indicator variable for whether the previous period return falls into the top or bottom CAR quintile and (ii) Q5 Return ($\mathbb{1}_{i,t}^{Q5}$), an indicator variable for whether the previous period return falls into the top CAR quintile. Control variables include price, idiosyncratic volatility, idiosyncratic skewness, log size, book-to-market, momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. Independent variables are winsorized and standardized. Regressions are split by large cap and small cap samples, where large cap includes all stocks with a market cap over $10bn and small cap includes all stocks with a market cap less than or equal to $2bn. Both regressions include time fixed effects and have standard errors that are clustered by time and security.

Panel A: Small Cap

	% ΔRH Investors
Q1 or Q5 return ($\mathbb{1}_{i,t}^{Q1or5}$)	0.0049***
	(12.23)
Q5 return ($\mathbb{1}_{i,t}^{Q5}$)	0.0047***
	(7.60)
Controls	Yes
R-squared	0.01
Time fixed effects	Yes
Num. of observations	746,505

Panel B: Large Cap

	% ΔRH Investors
Q1 or Q5 return ($\mathbb{1}_{i,t}^{Q1or5}$)	0.0103***
	(14.32)
Q5 return ($\mathbb{1}_{i,t}^{Q5}$)	-0.0081***
	(-12.46)
Controls	Yes
R-squared	0.04
Time fixed effects	Yes
Num. of observations	213,404

Table 8

Buy-the-Dip Effect, Regression Analysis

This table reports the regression results demonstrating the buy-the-dip effect for large-cap companies. The dependent variable is the percent change in the number of Robinhood investors. The control variables include price, idiosyncratic volatility, idiosyncratic skewness, log size, book-to-market, momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. Top 100 is an indicator variable set to one for the 100 most popular stocks on the Robinhood platform by number of investors at each point in time and zero otherwise. Independent variables are winsorized and standardized. The regression includes time fixed effects and standard errors are clustered by time and security.

	1-Day % Change in Robinhood Investors		
	(1)	(2)	(3)
Previous day return (abs. value)	0.0038***	0.0038***	0.0037***
	(17.09)	(17.09)	(16.79)
Previous day return	-0.0007***	-0.0016***	-0.0017***
	(-3.21)	(-4.22)	(-4.45)
Previous day return * $\mathbb{1}^{LC}$		-0.0001**	-0.0001**
		(-2.65)	(-2.39)
Previous day return * $\mathbb{1}^{SC}$		0.0010***	0.0010***
		(4.15)	(4.16)
Previous day return * $\mathbb{1}^{LC}$ * Top 100			-0.0001**
			(-2.76)
Controls	Yes	Yes	Yes
R-squared	3.9%	3.9%	3.9%
Time fixed effects	True	True	True
Number of observations	1,486,305	1,486,305	1,486,305

Table 9
Trading in Response to Volume and Extreme Returns

This table reports panel regression estimates for the percent change in the number of Robinhood investors on stock characteristics. The dependent variable in these regressions is the one-day percent change in the number of Robinhood investors for stock i and time t. All independent variables are measured at the end of the previous day. The control variables include momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. All independent variables are standardized. Regressions include time fixed effects and standard errors are clustered by time and security.

	1-Day % Change in Robinhood Investors		
	(1)	(2)	(3)
Previous day return	-0.0006***		-0.0007***
	(-3.51)		(-3.20)
Previous day return (abs. value)	0.0036***		0.0038***
	(16.14)		(16.30)
Dollar volume traded		0.0001	0.0004***
		(0.37)	(8.75)
Abnormal dollar volume traded		0.0002	0.0007
		(0.91)	(0.44)
Controls	No	No	Yes
R-squared	0.03	0.02	0.04
Time fixed effects	True	True	True
Number of observations	1,763,397	1,654,688	1,345,056

Table 10

Robinhood Investment and Earnings Surprises

This table shows the average three-day percent change in the number of Robinhood investors for stocks in a given standardized unexpected earnings (SUE) quintile. Quintile 1 represents the lowest SUE quintile and Quintile 5 represents the highest SUE quintile. The control group includes all observations that did not have an earnings announcement. Small Cap includes all observations for stocks with a market cap less than or equal to $2bn, Mid Cap includes all observations for stocks with a market cap between $2bn and $10bn, and Large Cap includes all observations for stocks with a market cap over $10bn. The Overall column includes observations across all market caps. The Observation Count column reports the number of observations used to compute the quintile averages in the Overall column.

SUE Quintile	Small Cap	Mid Cap	Large Cap	Overall	Observation Count
Quintile 1	6.4%	6.3%	4.6%	6.2%	5,439
Quintile 2	5.3%	5.9%	4.0%	5.4%	5,740
Quintile 3	4.7%	5.6%	4.5%	5.0%	5,578
Quintile 4	5.2%	5.8%	4.8%	5.3%	5,518
Quintile 5	6.3%	7.6%	5.3%	6.6%	4,834
Control	1.3%	0.8%	0.7%	1.1%	1,767,306

Table 11

Robinhood Investment and Analyst Recommendation Revisions

This table shows the average three-day percent change in the number of Robinhood investors for stocks in a given analyst recommendation revision category. Rating downgrade includes all 3-day period observations when a stock experienced a net rating downgrade, and rating upgrade includes all 3-day period observations when a stock experienced a net rating upgrade. The control group includes all 3-day period observations that did not include a rating change. Analyst recommendations data comes from IBES. Small Cap includes all observations for stocks with a market cap less than or equal to $2bn, Mid Cap includes all observations for stocks with a market cap between $2bn and $10bn, and Large Cap includes all observations for stocks with a market cap over $10bn. The Overall column includes observations across all market caps. The Observation Count column reports the number of observations used to compute the averages in the Overall column.

Rating Change	Small Cap	Mid Cap	Large Cap	Overall	Observation Count
Rating downgrade	17.2%	7.6%	4.0%	8.8%	6,459
Rating upgrade	9.2%	3.6%	2.2%	4.0%	5,651
Control	1.4%	0.8%	0.8%	1.1%	1,782,233

Table 12

The Effect of WallStreetBets Popularity on Robinhood Investment

This table reports the results of a regression of the percent change in the number of Robinhood investors on four WallStreetBets variables: (i) post mention indicator, indicating whether the stock was mentioned on the WallStreetBets platform at all on the previous day; (ii) excess winsorized comments, the excess and winsorized value of the sum of associated comments on the previous day; (iii) excess winsorized score, the excess and winsorized value of the sum of the associated scores on the previous day; and (iv) awards, the sum of awards for posts that mention the stock on the previous day. The control variables include price, idiosyncratic volatility, idiosyncratic skewness, log size, book-to-market, momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. Independent variables are winsorized and standardized. The regression includes time fixed effects and standard errors are clustered by time and security.

	1-Day % Change in Robinhood Investors				
	(1)	(2)	(3)	(4)	(5)
Post mention indicator	0.0004***				0.0004***
	(7.69)				(5.81)
Excess winsorized comments		0.0004***			0.0004***
		(8.00)			(4.78)
Excess winsorized score			0.0003***		-0.0002***
			(6.84)		(-3.10)
Awards				0.0001**	0.0001
				(2.24)	(1.22)
Controls	Yes	Yes	Yes	Yes	Yes
R-squared	4.20%	4.19%	4.18%	4.16%	4.21%
Time fixed effects	Yes	Yes	Yes	Yes	Yes
Number of observations	1,121,664	1,121,663	1,121,663	1,121,663	1,121,663

Table 13
WallStreetBets Sentiment Dictionary

This table reports the top phrases that indicate bullish sentiment and the top ten phrases that indicate bearish sentiment on the WallStreetBets platform.

Bullish	Bearish
Buy	Sell
Call	Put
Bull	Bear
Long	Short
YOLO	Get out
All in	Crash
To the moon	Down
Beat earnings	Miss earnings
Undervalued	Overvalued
🚀	Overbought
💎 🙌	🐻

Table 14
The Effect of WallStreetBets Sentiment on Robinhood Investment

This table reports the results of a panel regression of the percent change in the number of Robinhood investors on a measure of WallStreetBets sentiment. The sentiment variable is constructed as outlined in Section 4.3, and reflects WallStreetBets platform sentiment for each date-ticker observation. The control variables include price, idiosyncratic volatility, idiosyncratic skewness, log size, book-to-market, momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. Independent variables are winsorized and standardized. The regression includes time fixed effects and standard errors are clustered by time and security.

	1-Day % Change in Robinhood Investors
Sentiment	0.0002***
	(4.00)
Controls	Yes
R-squared	4.17%
Time fixed effects	Yes
Number of observations	1,121,663

Table 15
Robinhood Representative Portfolio - Performance Statistics

This table reports the performance statistics of the Robinhood representative portfolio (dollar method), Robinhood representative portfolio (share method), and the Market portfolio. The Robinhood representative portfolio using the dollar method is constructed by assuming every Robinhood investor in a stock represents an equal dollar amount investment. The Robinhood representative portfolio using the share method is constructed by assuming every Robinhood investor in a stock represents an equal share amount investment. The market return is computed as the value-weighted return of all U.S. stocks. Due to the Robintrack sample date range, the 2018 period covers 05/03/2018 - 12/31/2018, the 2019 period covers 01/01/2019 - 12/31/2019, and the 2020 period covers 01/01/2020 - 08/14/2020.

Period	2018	2019	2020	Full Period	Full Period (ann.)	Standard Deviation (ann.)	Sharpe Ratio
Robinhood portfolio (dollar method)	-6.0%	38.3%	33.6%	73.7%	27.4%	30.9%	0.83
Robinhood portfolio (share method)	-6.3%	33.7%	45.7%	82.5%	30.1%	29.3%	0.97
Market portfolio	-4.8%	30.5%	7.4%	33.3%	13.4%	24.9%	0.47

Table 16

Robinhood Representative Portfolio - Model Alphas

This table reports the results of regressions of the return of the Robinhood aggregate representative portfolio constructed using the dollar method on the market return less the risk-free rate, the value factor HML, the size factor SML, the momentum factor MOM, the profitability factor RMW, and the investment factor CMA. The performance time period is the full sample period May 3, 2018 - August 14, 2020. Market and factor returns are sourced from Kenneth R. French's data library. t-statistics are Newey-West adjusted for two lags.

	CAPM	3 Factor Model	4 Factor Model	6 Factor Model
Intercept	0.0006	0.0008***	0.0006***	0.0007***
	(1.60)	(2.81)	(2.34)	(2.88)
Market	1.13***	1.06***	1.07***	1.03***
	(25.72)	(32.67)	(43.04)	(38.42)
HML		0.01	-0.26***	-0.11*
		(0.17)	(-4.85)	(-1.80)
SMB		0.67***	0.56***	0.45***
		(12.58)	(6.74)	(7.52)
MOM			-0.33***	-0.36***
			(-4.89)	(5.03)
RMW				-0.23***
				(-2.67)
CMA				-0.56***
				(-4.77)
R-squared	82.8%	89.3%	91.0%	92.1%
Number of obs.	566	566	566	566

Table 17

Strategy Performance for Buy-the-Dip and Event-Based Trading

This table reports the strategy performance for the buy-the-dip effect strategy and the event-based trading strategy. Panel A reports the results for the Buy-the-Dip strategy. The BTD row reports the average future 1-day returns and 3-day returns for stocks that are large cap and fall into the lowest return decile in the present day. The Non-BTD row reports the average future 1-day returns and 3-day returns for all other stocks. Panel B reports the results for the Event-Based Trading strategy. The EA row reports contemporaneous 1-day returns and 3-day returns for stocks that announce earnings on the present day and the Non-EA row reports contemporaneous 1-day returns and 3-day returns for all other stocks. The remaining rows in Panel B show the average 1-day and 3-day return performance broken out by SUE quintile.

Panel A - Buy-the-Dip

	May 2018 - August 2020		**May 2015 - June 2023**	
	3-day return (forward)	1-day return (forward)	3-day return (forward)	1-day return (forward)
BTD	-6.0483%	0.1060%	-6.2834%	0.4891%
Non-BTD	0.1426%	0.0422%	0.1810%	0.0540%

Panel B - Event-Based Trading

	May 2018 - August 2020		**May 2015 - June 2023**	
	3-day return	1-day return	3-day return	1-day return
Sue_q1	-0.0008%	-0.4562%	-0.0249%	-0.4677%
Sue_q2	-0.0271%	-0.3670%	-0.0407%	-0.3710%
Sue_q3	0.5116%	0.2934%	0.4468%	0.2532%
Sue_q4	0.9660%	0.6280%	0.9273%	0.6013%
Sue_q5	1.1570%	0.8445%	1.1490%	0.8345%
EA	0.5112%	0.1795%	0.4876%	0.1681%
Non-EA	0.1440%	0.0547%	0.1084%	0.0404%

Table 18
Robinhood Investment and Return Predictability

This table reports the return predictability regression results for weekly returns. The dependent variables in columns (1), (2), (3), and (4) are the following week's return, return two weeks out in the future, return three weeks out in the future, and return four weeks out in the future. NI_t^i is the log winsorized number of Robinhood investors in stock i in the previous period. $ret_{prev}^{i,d}$, $ret_{prev}^{i,w}$, and $ret_{prev}^{i,m}$ are the previous daily, weekly, and monthly returns respectively. $p212_{prev}^i$ is the value of the previous year's return less the past month return. All independent variables are measured at the end of the previous day. The control variables include momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. All independent variables are standardized. Regressions include time fixed effects and standard errors are clustered by time and security.

	Future Weekly Return			
	(1)	(2)	(3)	(4)
NI_{prev}^i	0.14**	0.13**	0.13**	0.13**
	(2.37)	(2.28)	(2.30)	(2.28)
$ret_{prev}^{i,d}$	-0.06***	0.17	1.32	-2.02
	(-3.36)	(0.09)	(0.53)	(-1.03)
$ret_{prev}^{i,w}$	-0.02	-0.84	-0.83	-0.38
	(-1.61)	(-0.66)	(-0.64)	(-0.30)
$ret_{prev}^{i,m}$	-0.01	-0.43	-0.17	-0.05
	(-1.03)	(-0.77)	(-0.29)	(-0.08)
$p212_{prev}^i$	-0.07	-0.06	-0.07	-0.09
	(-0.38)	(-0.35)	(-0.42)	(-0.54)
Controls	Yes	Yes	Yes	Yes
R-squared	0.12	0.08	0.18	0.10
Time fixed effects	True	True	True	True
Number of observations	328,828	325,317	321,809	318,302

Figures

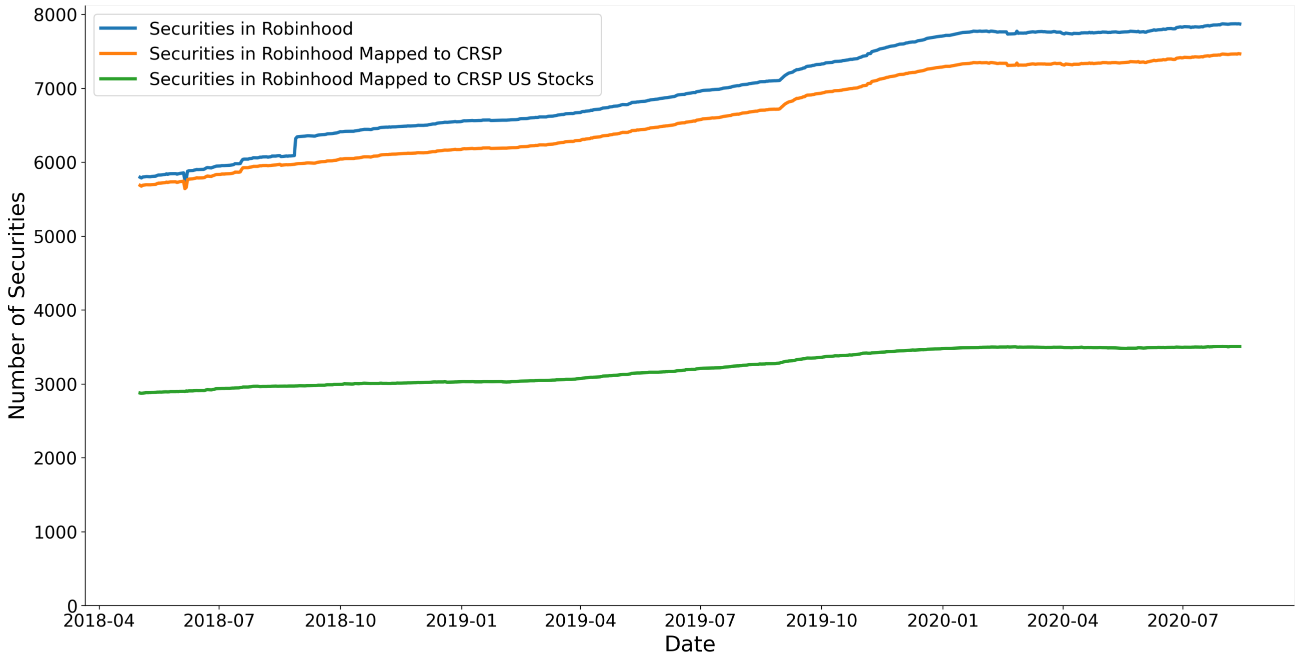


Figure 1: Robinhood Securities Matched to CRSP
This figure displays the coverage of Robinhood securities in CRSP over the sample period May 2, 2018 - August 13, 2020. Calculations of the number of securities are based on Robinhood tickers. The blue line represents the total number of Robinhood securities which had a positive number of Robinhood investors over time. The orange line represents the time series of the total number of Robinhood securities which both (i) had a positive number of Robinhood investors and (ii) were successfully mapped to a CRSP permno. The green line represents the time series of the total number of Robinhood securities which (i) had a positive number of Robinhood investors; (ii) were successfully mapped to a CRSP permno; and (iii) were U.S. stocks, as denoted by share codes 10 and 11 in CRSP. The data comes from Robintrack and CRSP.

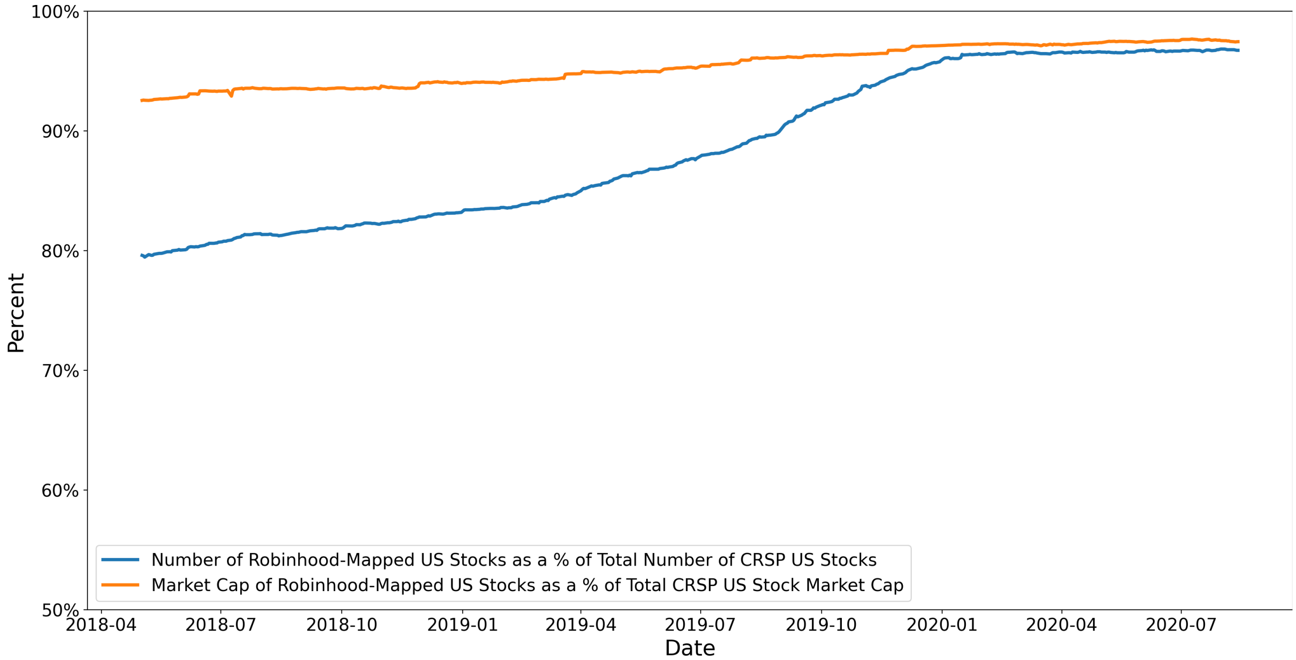
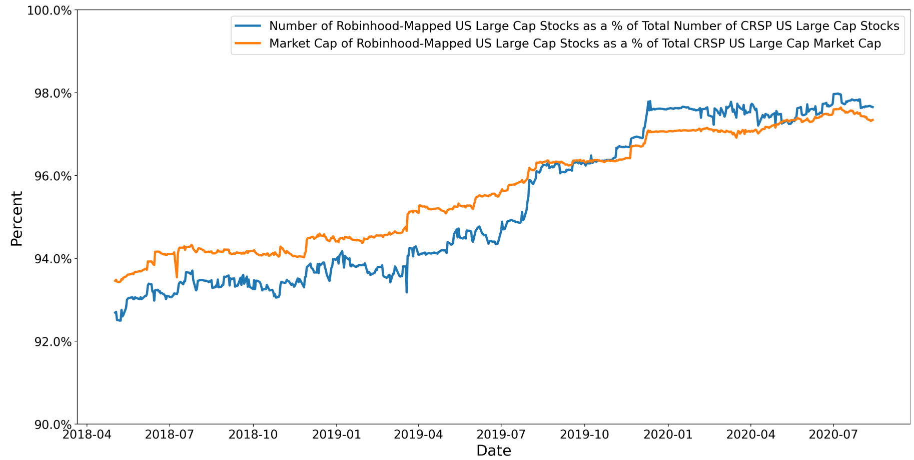
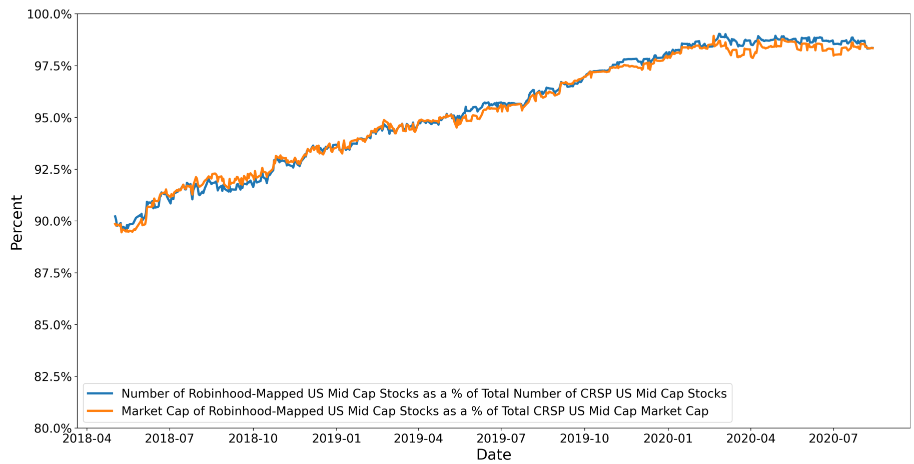


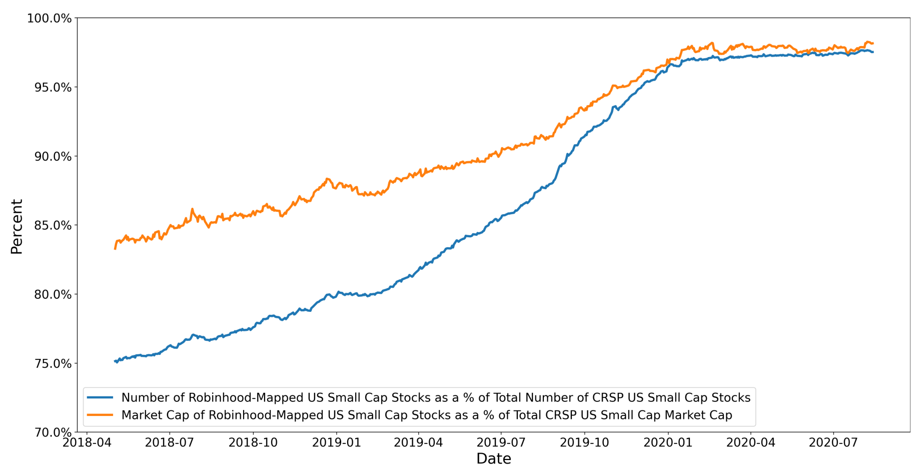
Figure 2: CRSP U.S. Stock Securities Matched to Robinhood
This figure displays the coverage of CRSP U.S. stock securities in Robinhood over the sample period May 2, 2018 - August 13, 2020. The blue line represents the time series of the percent of U.S. stocks in CRSP that can be mapped to Robinhood securities. The orange line represents the time series of the market cap of U.S. stocks that can be mapped to Robinhood as a percent of the total market cap of all CRSP U.S. stocks. The data comes from Robintrack and CRSP.



(a) Large Cap



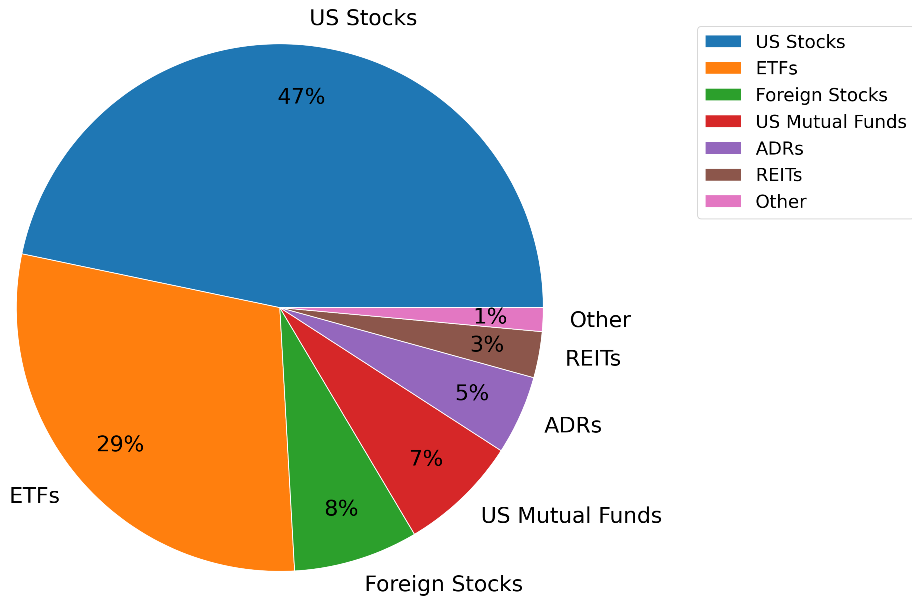
(b) Mid Cap



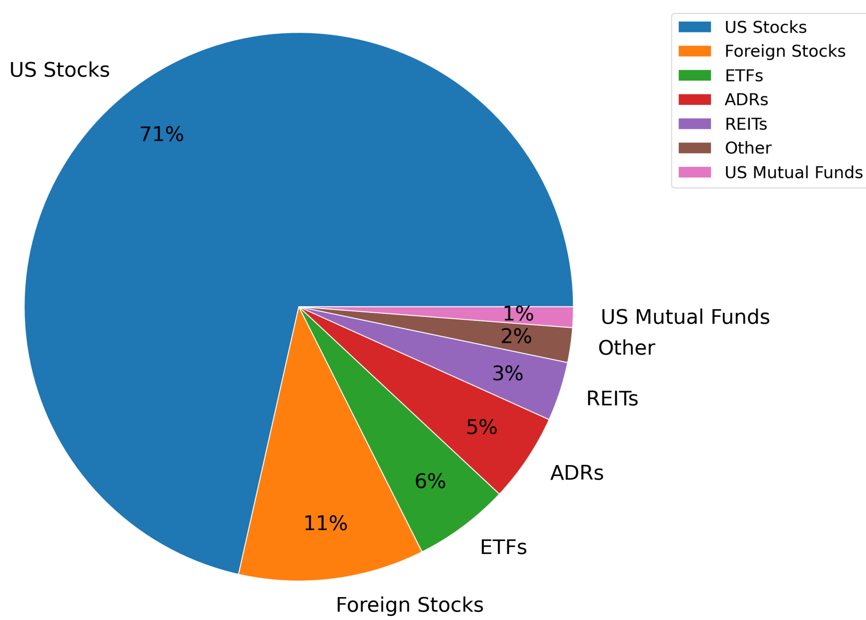
(c) Small Cap

Figure 3: CRSP U.S. Stock Securities Matched to Robinhood by Market Cap

This figure displays the coverage of CRSP U.S. stock securities in Robinhood by market cap over the sample period May 2, 2018 - August 13, 2020. The blue line represents the time series of the percent of U.S. stocks in CRSP in a given size group that can be mapped to Robinhood securities. The orange line represents the time series of the market cap of U.S. stocks in a given size group that can be mapped to Robinhood as a percent of the total market cap of all CRSP U.S. stocks in that size group. The data comes from Robintrack and CRSP. Panel (a) shows the results for large-cap stocks, where large cap is defined as a market cap of at least \$10bn. Panel (b) shows the results for mid-cap stocks, where mid cap is defined as a market cap between \$2bn and \$10bn. Panel (c) shows the results for small-cap stocks, where small cap is defined as a market cap less than or equal to \$2bn.



(a) Number of Securities Breakdown



(b) Number of Investors Breakdown

Figure 4: Number of Robinhood Securities and Investors by Security Type

The pie chart in Panel (a) displays the percent of Robinhood securities that were mapped to CRSP and fall under each of the following security types: U.S. Stocks, ETFs, Foreign Stocks, U.S. Mutual Funds, ADRs, REITs, and Other. The pie chart in Panel (b) first sums the number of Robinhood investors in each security type, and then displays that as a percent of the total number of Robinhood investors. Robinhood securities are classified into the following security types: U.S. Stocks, Foreign Stocks, ETFs, ADRs, REITs, U.S. Mutual Funds, and Other. The security type is determined by the share code attribute from CRSP. The total number of securities used in the analysis is the 7,969 unique CRSP permnos to which the Robintrack securities were mapped. For each CRSP permno which corresponded to more than one share code over its lifetime, the most recent date share code mapping was used. The number of investors and security type data is taken from August 13, 2020, the latest available date of Robintrack data.

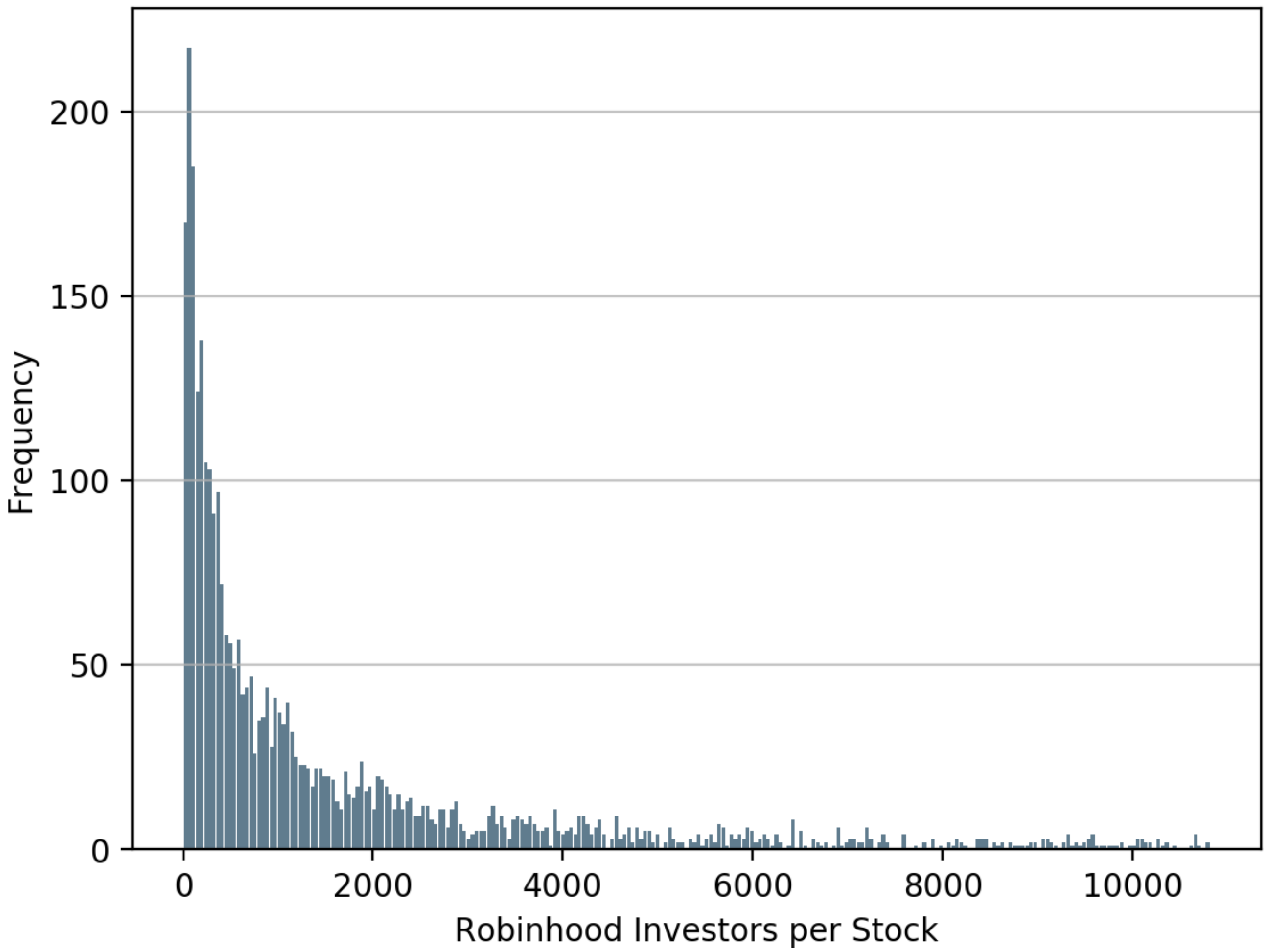


Figure 5: Histogram of the Number of Robinhood Investors per Stock
At each point in time, each stock in the Robintrack dataset is held by a certain number of investors. This histogram depicts the frequency with which stocks had a certain number of investors. In order to improve visibility, the right tail of the distribution is truncated at the 90th percentile. This analysis relies on data from August 13, 2020, the latest available date in the Robintrack data.

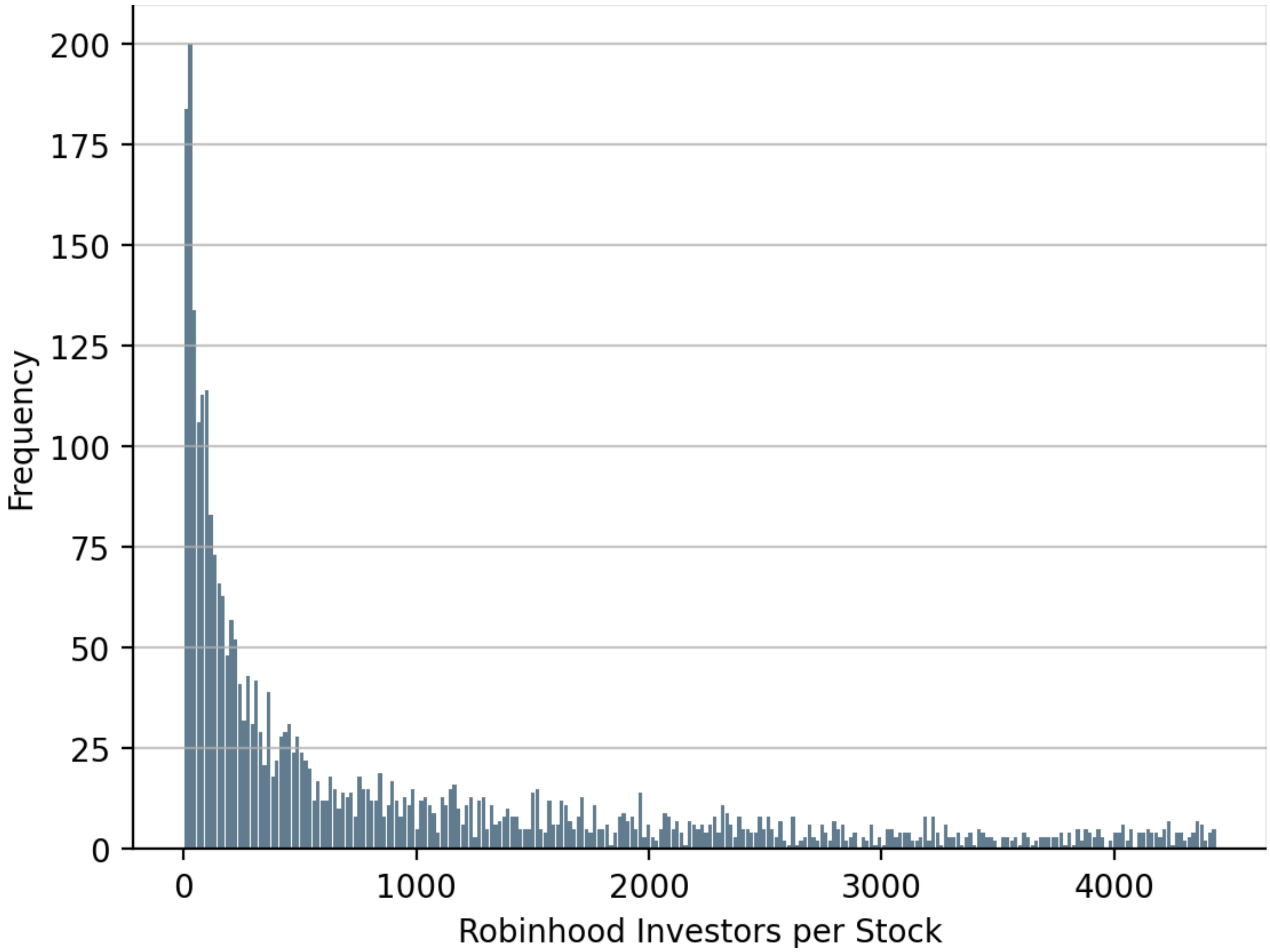
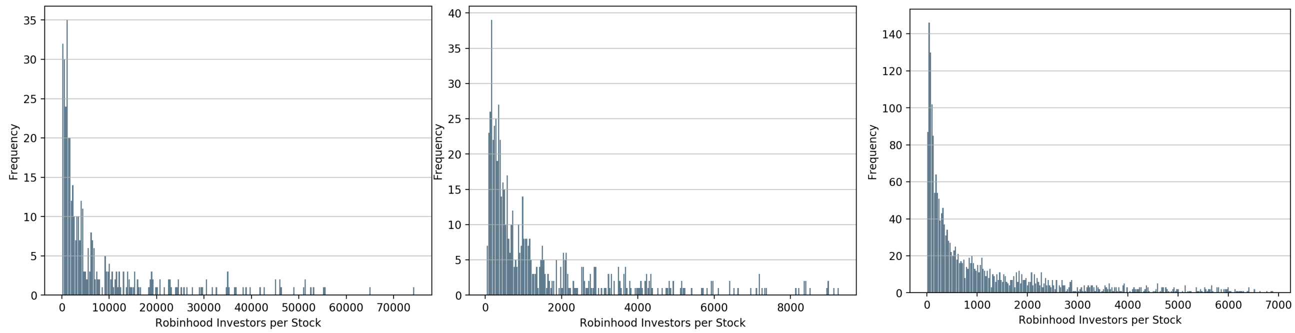


Figure 6: Simulated Histogram of the Number of Robinhood Investors per Stock
This simulated histogram depicts the frequency with which stocks had a certain number of investors. The parameters of the simulation include 5,000 investors, a stock universe containing approximately 3,500 stocks, and a portfolio of ten stocks per investor. The probability that an investor chooses a given stock as one of the ten stocks in their portfolio is equal to the stock's market weight. In order to improve visibility, the right tail of the distribution is truncated at the 90th percentile.



(a) Large Cap **(b) Mid Cap** **(c) Small Cap**

Figure 7: Histogram of the Number of Robinhood Investors per Stock, by Market Cap
These histograms depict the frequency with which stocks had a certain number of Robinhood investors. In order to improve visibility, the right tails of the distributions are truncated at the 90th percentile. This analysis relies on data from August 13, 2020, the latest available date in the Robintrack data. Panel (a) shows the results for large-cap stocks, where large cap is defined as a market cap of at least \$10bn. Panel (b) shows the results for mid-cap stocks, where mid cap is defined as a market cap between \$2bn and \$10bn. Panel (c) shows the results for small-cap stocks, where small cap is defined as a market cap less than or equal to \$2bn.

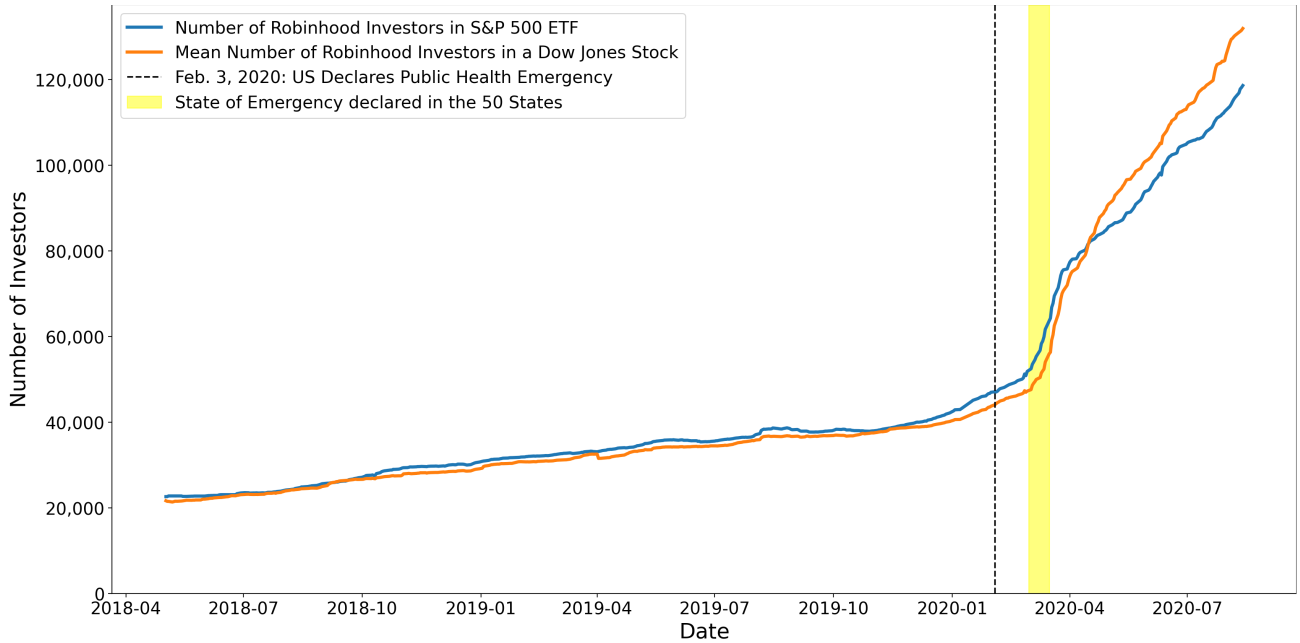
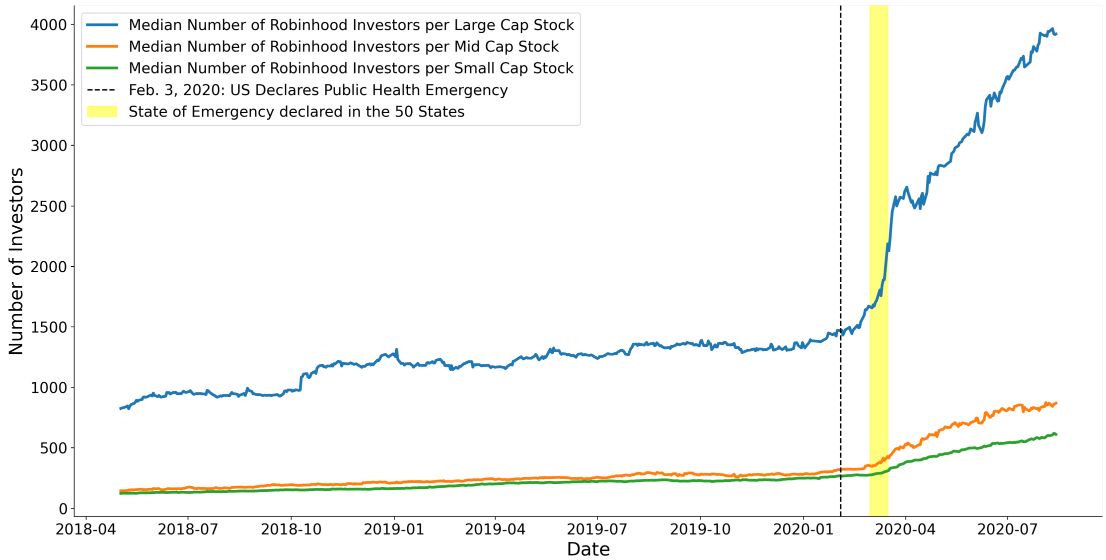
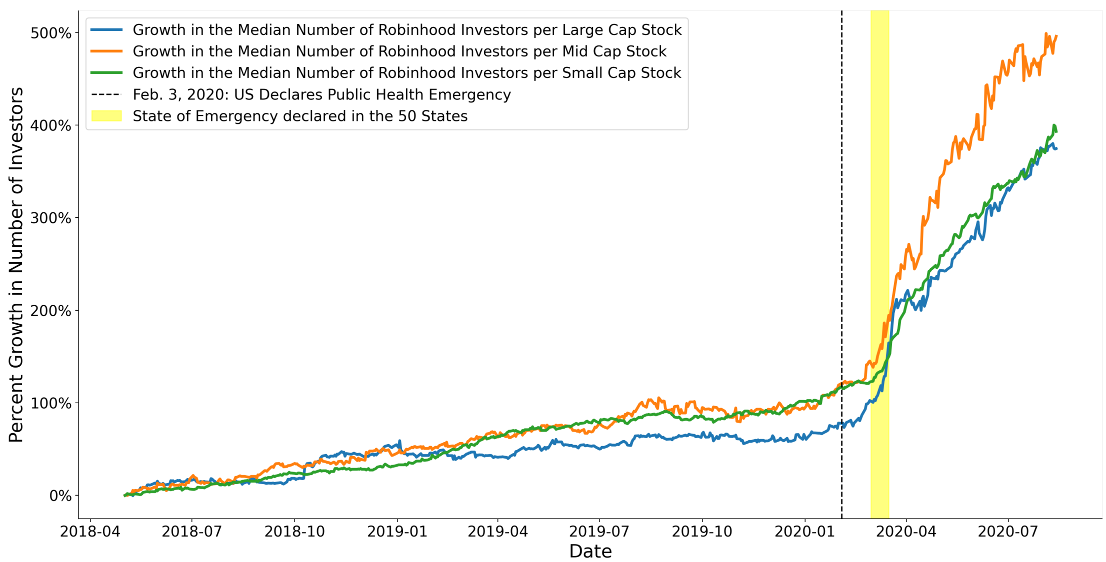


Figure 8: Pandemic Shock to Robinhood Ownership

This figure shows that the start of the Covid-19 pandemic coincided with a significant increase in the number of Robinhood investors. The blue line represents the number of Robinhood investors in the S&P 500 ETF over time. The orange line represents the average number of Robinhood investors invested in one of the 30 Dow Jones Industrial Average (DJIA) stocks over time. The dashed line represents February 3, 2020, which is the day that the U.S. declared Covid-19 to be a public health emergency. The yellow shaded area represents the range of time when the 50 U.S. states announced their states of emergency, with the beginning of the yellow shaded area corresponding to the earliest state of emergency declaration in Washington on February 29, 2020 and the end of the yellow shaded area corresponding to the latest state of emergency declaration in Vermont on March 16, 2020.



(a) Raw Number



(b) Percent Growth

Figure 9: Pandemic Shock to Robinhood Ownership, by Market Cap

This figure displays the growth of the median number of Robinhood investors per stock over the sample period May 2, 2018 - August 13, 2020, and shows that the impact of the Covid-19 pandemic was not limited to stocks of a particular size. The blue line represents large cap stocks, where large-cap is defined as a market cap of at least $10bn. The orange line represents mid-cap stocks, where mid cap is defined as a market cap between $2bn and $10bn. The green line represents small-cap stocks, where small cap is defined as a market cap less than or equal to $2bn. Panel (a) shows the raw median number of Robinhood investors per stock for each market cap group over time. Panel (b) shows the cumulative percent growth in the median number of Robinhood investors per stock for each market cap group over time.

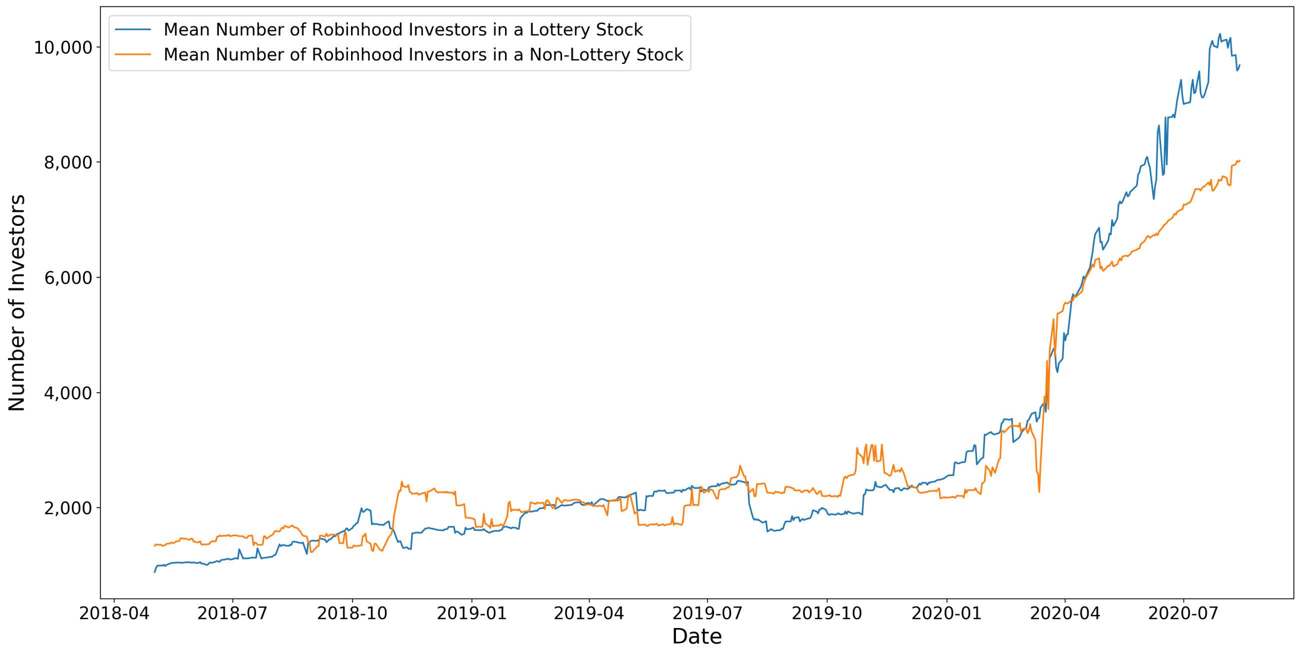


Figure 10: Mean Robinhood Investors in Lottery Stocks and Non-Lottery Stocks
This chart plots a time series of the average number of Robinhood investors in a lottery stock and in a non-lottery stock over time. Lottery stocks are defined as those stocks whose price falls into the lowest 50th percentile, whose idiosyncratic volatility falls into the highest 50th percentile, and whose idiosyncratic skewness falls into the highest 50th percentile. Non-lottery stocks are defined as those stocks whose price falls into the highest 50th percentile, whose idiosyncratic volatility falls into the lowest 50th percentile, and whose idiosyncratic skewness falls into the lowest 50th percentile.

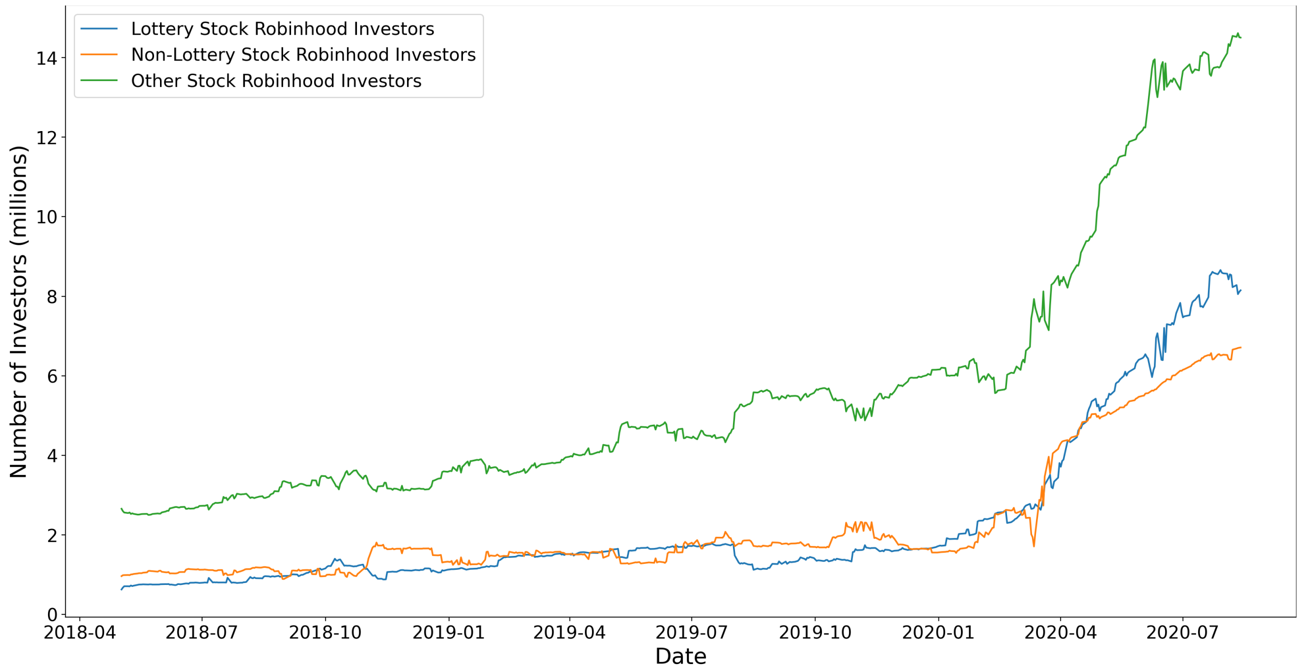


Figure 11: Aggregate Robinhood Investors in Lottery Stocks, Non-Lottery Stocks, and Other Stocks

This chart plots a time series of the total number of Robinhood investors across all lottery stocks, non-lottery stocks, and other stocks over time. Lottery stocks are defined as those stocks whose price falls into the lowest 50th percentile, whose idiosyncratic volatility falls into the highest 50th percentile, and whose idiosyncratic skewness falls into the highest 50th percentile. Non-lottery stocks are defined as those stocks whose price falls into the highest 50th percentile, whose idiosyncratic volatility falls into the lowest 50th percentile, and whose idiosyncratic skewness falls into the lowest 50th percentile. Other stocks include all remaining stocks.

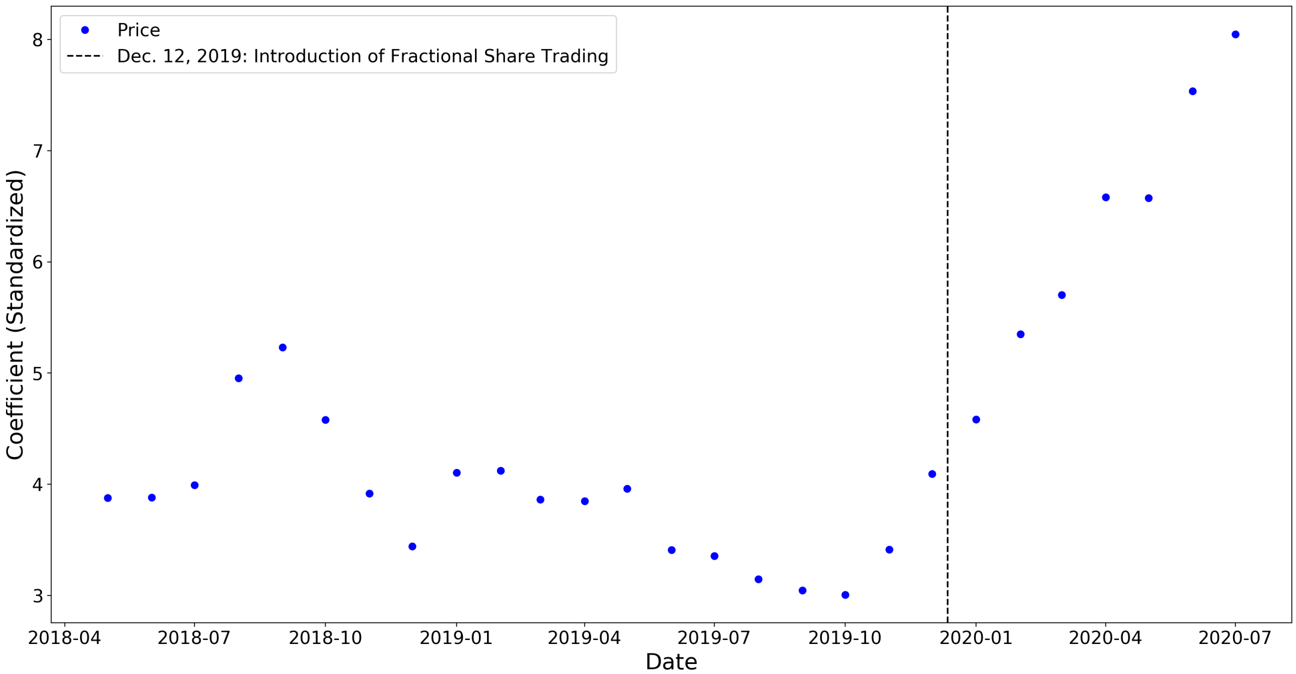
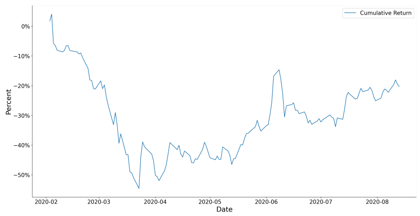
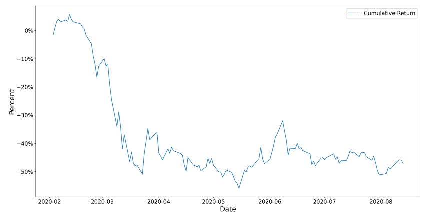


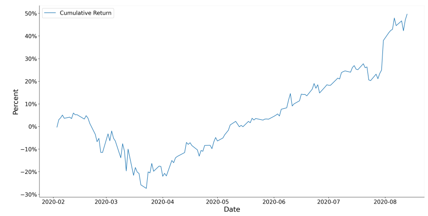
Figure 12: Price Coefficient in the Number of Robinhood Investors Regression
This chart plots the time series of the coefficient on price over monthly panel regressions of the number of Robinhood investors on previous day stock price, idiosyncratic volatility, and idiosyncratic skewness. All panel regressions include time fixed effects, and standard errors are clustered by time and security. The chart shows that the coefficient on price in the regression increased following the introduction of fractional share trading on the Robinhood platform in December 2019.



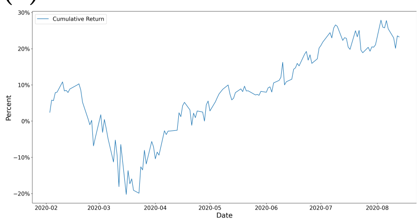
(a) Ford



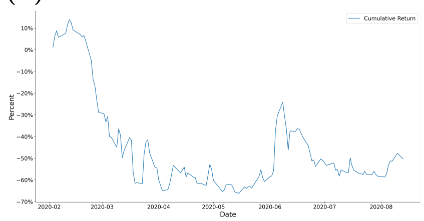
(b) General Electric



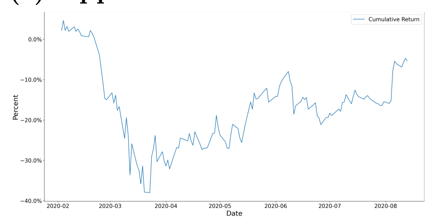
(c) Apple



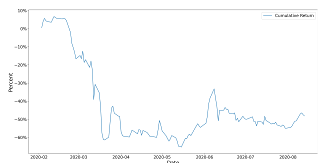
(d) Microsoft



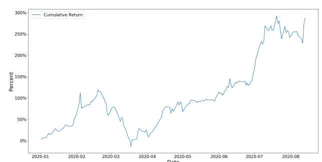
(e) American Airlines



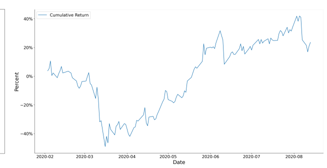
(f) Disney



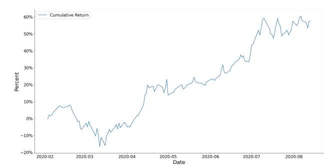
(g) Delta Airlines



(h) Tesla



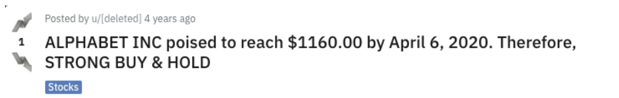
(i) GoPro



(j) Amazon

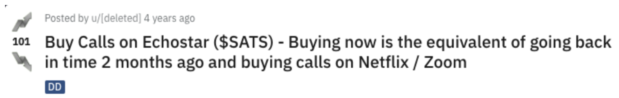
Figure 13: Performance of Top 10 Robinhood Holdings

The ten charts in this figure show the cumulative performance of the top 10 U.S. stock holdings of Robinhood investors over the period May 2, 2018 - August 13, 2020. The top U.S. stock holdings of Robinhood investors are determined as of August 13, 2020, the sample end date. Six of the top 10 Robinhood holdings experienced a drawdown of over 50% since the beginning of the sample period. Investing in large, well-known companies that fell upon hard times is in line with the buy-the-dip effect.



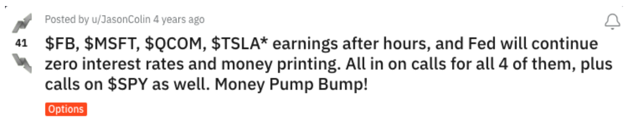
(a) **Bullish post #1**

(b) **Bearish post #1**

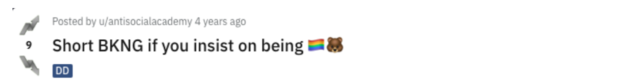


(c) **Bullish post #2**

(d) **Bearish post #2**



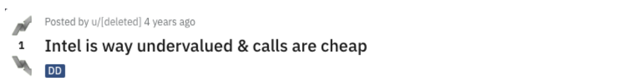
(e) **Bullish post #3**



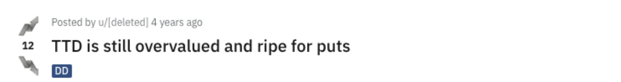
(f) **Bearish post #3**

(g) **Bullish post #4**

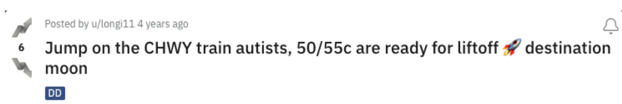
(h) **Bearish post #4**



(i) **Bullish post #5**



(j) **Bearish post #5**



(k) **Bullish post #6**

(l) **Bearish post #6**

Figure 14: WallStreetBets Sample Bullish and Bearish Posts
This figure displays six bullish posts and six bearish posts from the WallStreetBets platform. The posts appeared on the WallStreetBets platform over 2020, and the concatenations, slang terms, and emojis found therein are representative of the new-age terminology frequently adopted on the platform.

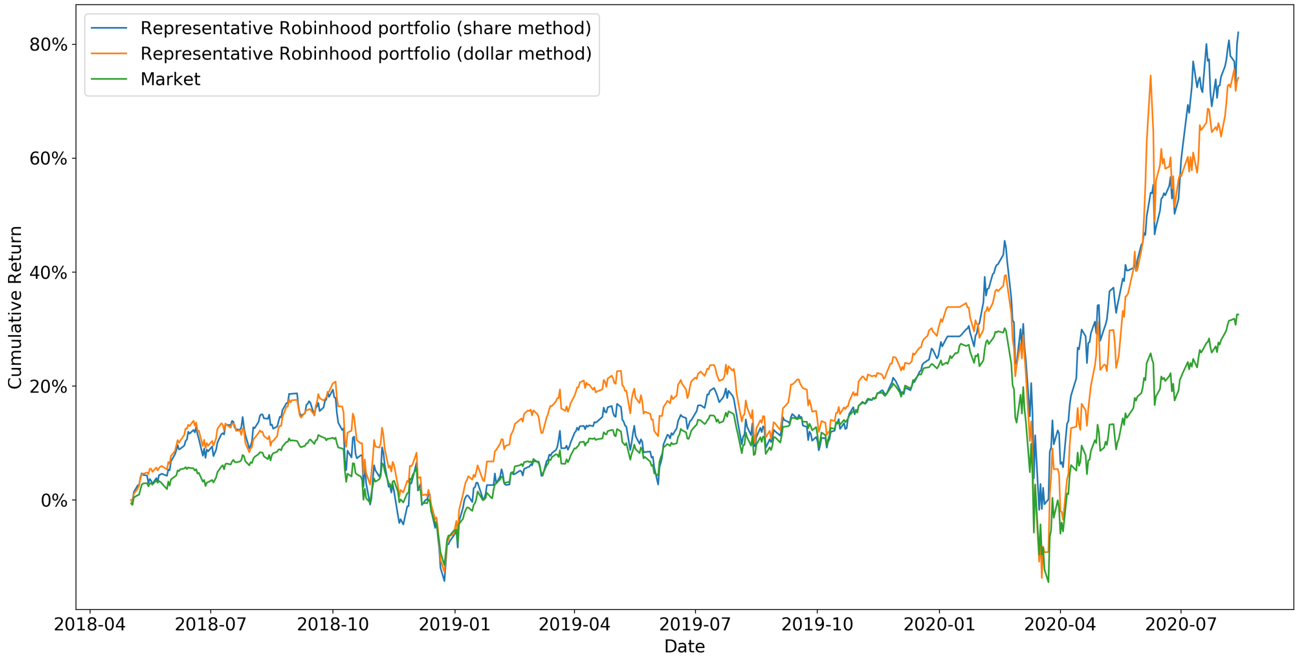
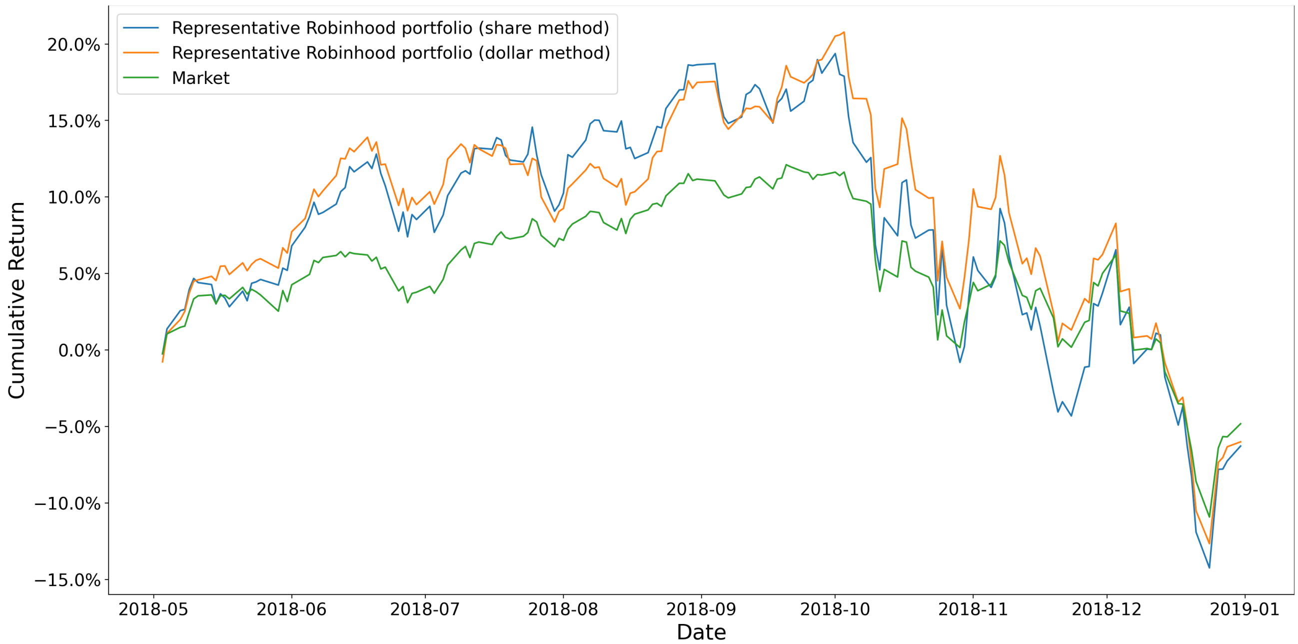
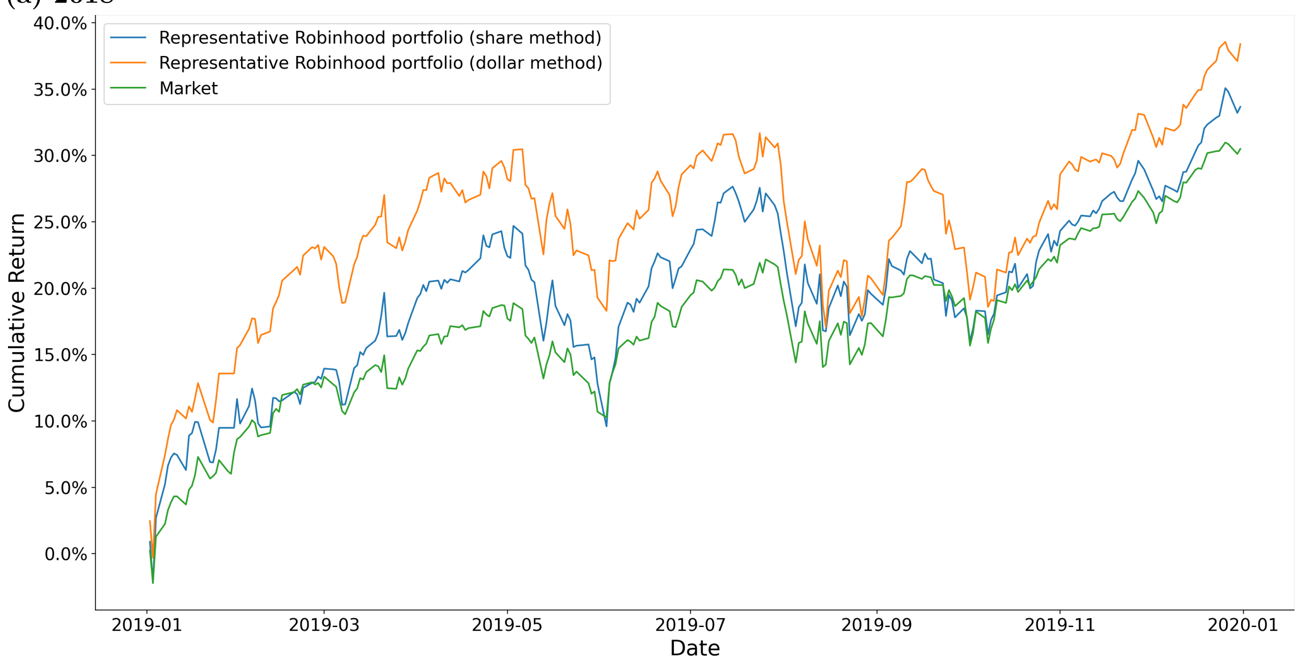


Figure 15: Robinhood Representative Portfolio Performance
This chart shows the cumulative performance of the Robinhood representative portfolios using the dollar method and the share method, and the market portfolio over the sample period May 2, 2018 - August 13, 2020. In the dollar method, each investor represents a $1 investment. In the share method, each investor represents a 1 share investment. Market returns are computed as the value-weighted return on all U.S. stocks.



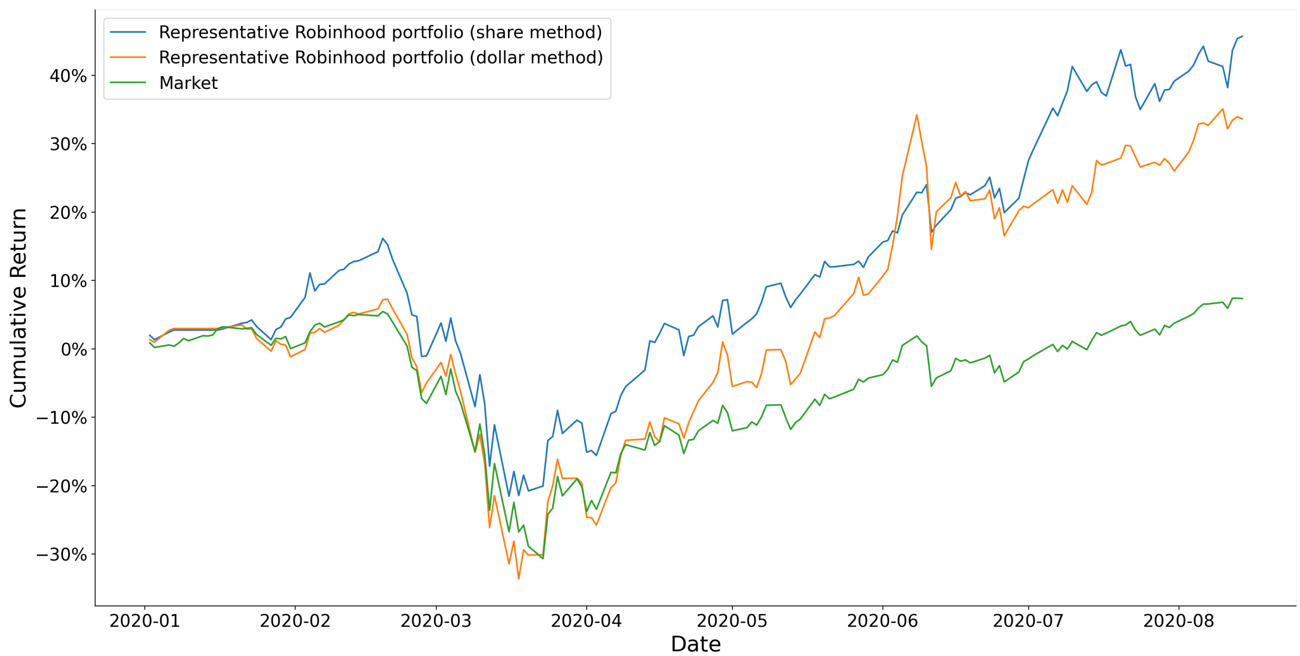
Representative Robinhood portfolio (share method)
Representative Robinhood portfolio (dollar method)
Market
Cumulative Return
20.0%
15.0%
10.0%
5.0%
0.0%
−5.0%
−10.0%
−15.0%
2018-05
2018-06
2018-07
2018-08
2018-09
2018-10
2018-11
2018-12
2019-01
Date

(a) 2018



Representative Robinhood portfolio (share method)
Representative Robinhood portfolio (dollar method)
Market
Cumulative Return
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
2019-01
2019-03
2019-05
2019-07
2019-09
2019-11
2020-01
Date

(b) 2019



(c) 2020

Figure 16: Robinhood Representative Portfolio Performance by Year

This chart shows the cumulative performance of the Robinhood representative portfolios using the dollar method and the share method, and the market portfolio over the sample period May 3, 2018 - December 31, 2018 (Panel A), January 1, 2019 - December 31, 2019 (Panel B), and January 1, 2020 - August 14, 2020 (Panel C). In the dollar method, each investor represents a $1 investment. In the share method, each investor represents a 1 share investment. Market returns are computed as the value-weighted return on all U.S. stocks.

Internet Appendix for:
This Time is Different: Investing in the Age of Robinhood

A Variable Descriptions

- **Market Capitalization:** Calculated as the number of shares outstanding ("shrout") multiplied by the stock price (absolute value of "prc"). Data used in the calculations is from CRSP. Stock price is the absolute value of "prc" as "prc" is set to the negative of the bid/ask average in CRSP when the closing price is not available.

- **Book to Market Ratio:** Calculated as book value of equity divided by the total market value of equity. For firms in July of year *t* to June of year *t+1*, book values are for the fiscal year ending in calendar year *t-1* and market values are from the end of calendar year *t-1*. For firms with multiple share classes (identified as CRSP PERMCOs with multiple PERMNOs), market equity is combined for all share classes. Book value is calculated using Compustat data following Davis, Fama, and French (2000). Book value of equity is stockholders' equity plus deferred taxes and investment tax credit (if available) minus the book value of preferred stock. Depending on availability, stockholders' equity is measured in order of precedence as either stockholders' equity ("seq"), common equity ("ceq") plus par value of preferred stock ("pstk"), or total assets ("at") minus total liabilities ("lt"). Deferred taxes and investment tax credit is "txditc." Depending on availability, preferred stock is measured in order of precedence as the redemption value ("pstkrv"), the liquidation value ("pstkl"), or the par value ("pstk") of preferred stock.

- **Momentum:** Calculated as the return in the past year excluding the last month to eliminate high-frequency reversal, i.e. the 12-month return less the prior month return following Jegadeesh and Titman (1993) and using returns data from CRSP.

- **Gross Profitability:** Calculated using Compustat data following Novy-Marx (2013). For firms in July of year *t* to June of year *t+1*, data from the fiscal year ending in calendar year *t-1* is used. Calculated as gross profits (revenue ("revt") minus cost of goods sold ("cogs")) divided by total assets ("at"). Total assets are required to be greater than zero.

- **Investment:** Calculated using Compustat data following Cooper, Gulen, and Schill (2008). For firms in July of year *t* to June of year *t+1*, investment is calculated using total assets ("at") from the fiscal year ending in year *t-1* divided by total assets from the fiscal year ending in year *t-2*, minus 1. Total assets from year *t-2* and total assets from year *t-1* are required to be greater than zero.

- **Firm Age:** Calculated as the number of years from founding date to the current date. Firm founding dates are sourced from the Field-Ritter dataset of company founding dates, as used in Field and Karpoff (2002) and Loughran and Ritter (2004).

- **Amihud Illiquidity:** Calculated as the annual average of the daily ratio of absolute stock return to dollar trading volume following Amihud (2002). Data used in the calculations is from CRSP.

- **Standardized Unexpected Earnings (SUE):** Calculated as the reported quarterly earnings per share less the analyst forecast earnings for that quarter, divided by the standard deviation of the analyst forecasts that quarter.

- **Analyst Recommendation Revisions:** Calculated as the total number of upgrade recommendation revisions less the total number of downgrade recommendation revisions aggregated across analysts for a given stock-date pair. The underlying data for analyst recommendations is sourced from the Institutional Brokers' Estimate System (I/B/E/S) Recommendations - Detail database. Recommendations are scored on a scale of 1 to 5, and include {Strong Sell, Sell, Hold, Buy, Strong Buy}. For a given analyst-stock pair, a recommendation revision is defined as an event day on which the given analyst changes their recommendation for the stock from their previous recommendation. For the purposes of this computation, any move in the positive direction (an upgrade) is given a score of 1 and any move in the negative direction (a downgrade) is given a score of -1. For each stock-date pair, the aggregate analyst recommendation revision is then computed as the number of upgrades less the number of downgrades for the stock on that date across all analysts covering that stock.

B Data Appendix

The analysis in this paper relies on the following six main data sources: (1) Robintrack number of investors data; (2) CRSP security market data; (3) Compustat security fundamentals data; (4) Fama-French factor return data; (5) IBES analyst coverage and earnings data; and (6) WallStreetBets platform data. In this Appendix, I detail the data loading, cleaning, and merging procedures I use for the less standard Robintrack and WallStreetBets datasets, and provide descriptive statistics and coverage results.

B.1 Robintrack Data

I start with the Robintrack dataset, which is available for public download through Robintrack at https://robintrack.net/data-download. Note that the terms "Robintrack data" and "Robinhood data" are used in this paper interchangeably, and both are taken to mean all of the data on the number of Robinhood investors that has been made available through the Robintrack platform. This dataset covers observations for 8,597 unique securities from May 2, 2018 to August 13, 2020. The observations for each security include a series of timestamps and the corresponding numbers of Robinhood investors holding that security. The timestamps are expressed in UTC and generally have an hourly frequency.

For the purposes of this paper, I consider the Robintrack number of investors data at a daily frequency. To that effect, for each day for a given security I leave the last available observation of the number of investors for that security that occurred before 4pm Eastern Standard Time (EST) that day. After reducing the sample to these daily observations and merging the 8,597 security data files together into a single dataset, the Robintrack dataset contains 5,913,891 date-security observations. Note that when I merge the security data files together, I add a column for the security ticker (which is sourced from the filename of the security data file, e.g. ticker "AAPL" is sourced from the "AAPL.csv" data filename) to differentiate between the observations of various securities.

I also perform the following two data cleaning procedures on the tickers for each security:

1. First, I remove any leading underscore (_) characters in filenames when transforming them to tickers (e.g. for _PRN.csv and _OUT.csv), since Robintrack only includes those underscore characters in order to prevent built-in escape sequences from running.

2. Second, I create tsymbols from tickers by removing any periods (.) in filenames, which are used to separate ticker from share class (so the Robintrack filename "AKO.A" would correspond to the "AKOA" tsymbol; this is also how tickers and share classes are combined to form tsymbols in CRSP). I also perform error checking functions, such as ensuring that the number of datasets pulled in matches the original number of securities in Robintrack and ensuring that no security has duplicate number of investors data.

Next, I remove data for the 48 securities with multiple share classes for which the number of investors data are frequently not calculated properly in Robintrack. Details on these 48 securities and the types of associated data issues they experienced in Robintrack are outlined in Appendix B.3. Note that the exclusion of these 48 securities does not represent the comprehensive exclusion of all multiple share securities from the dataset; many remain in the dataset and do not exhibit any apparent data issues, a notable example being GOOG (Alphabet Inc Class C) and GOOGL (Alphabet Inc Class A). At this point I am left with 8,549 securities and 5,894,725 observations.

I then remove the observations for 08/14/2020 (which were formed as the latest available observations after 4pm on 08/13/2020), since they are not true end-of-date observations as the dataset terminated on 08/13/2020. There are 5,886,644 observations remaining after this step. I further remove securities with no positive number of investors data over the entire sample period. There are a total of 42 such securities. This leaves observations for 8,507 different securities.

For illustrative purposes, Table IA.1 reports the number of Robintrack observations and number of unique securities left after each step of the data cleaning procedure.

Table IA.1
Robintrack Securities and Observations Count

This table shows the number of unique securities and number of observations in the Robintrack dataset through a series of data cleaning steps. Observations refer to date-security observations, and unique securities are counted based on unique tickers as denoted in Robintrack.

Data cleaning step	Number of unique securities	Number of observations
Consolidating intraday to daily	8,597	5,913,891
Removing multiple share class securities	8,549	5,894,725
Removing 8/14/20 observations	8,549	5,886,644
Removing securities with no positive holdings	8,507	5,872,803

B.2 Ticker Symbol to PERMNO Mapping

The Robinhood dataset, following the data cleaning steps outlined in the previous sub-section, contains data for 8,507 securities each identified by a unique tsymbol. I add the fields first_posholding_date and last_posholding_date for each security to represent the first date and last date with a positive number of Robinhood investors in that security, respectively. The date range from first_posholding_date to last_posholding_date is then a proxy for the timeframe between 05/02/2018 and 08/13/2020 during which the security was actively traded on Robinhood. In Table IA.2 I provide an example of the first five securities in the Robinhood dataset and their corresponding values of first_posholding_date and last_posholding_date.

Table IA.2
Actively Traded Date Ranges for Five Securities

This table shows the first date and the last date with a positive number of Robinhood investors for the first five securities in the Robintrack dataset.

tsymbol	first_posholding_date	last_posholding_date
A	2018-05-02	2020-08-13
AA	2020-01-16	2020-08-13
AAAU	2018-08-16	2020-08-13
AACAY	2018-08-28	2020-08-13
AACG	2018-05-02	2020-08-13

I next use the crspq.dse quarterly update table to map each tsymbol to its PERMNO. PERMNO is the primary security identifier in CRSP, allowing for the tracing of securities over time. It is a unique and permanent security identifier, and unlike CUSIP, Ticker Symbol, and Company Name, a PERMNO does not change during an issue's trading history, nor is it reassigned after an issue ceases trading. This allows me to track the security through its entire trading history in CRSP files with a single PERMNO, regardless of name or capital structure changes. Stock data are sorted and indexed by this field, and the format of a PERMNO is currently a five-digit integer for all common securities in the CRSP files.

I first obtain a dataset of all the PERMNOs that were associated with the 8,507 Robinhood tsymbols at some point between the Robinhood sample start date (May 2, 2018) and end date (August 13, 2020).

To identify which of the associated PERMNOs is the correct PERMNO for each tsymbol, I proceed as follows. For each tsymbol-PERMNO pair from the crspq.dse dataset, I also obtain the start date and end date for when the tsymbol was assigned to that PERMNO. For each tsymbol in Robinhood, I identify the tsymbol-PERMNO pair in crspq.dse for which there is overlap between the [first positive holdings date, last positive holdings date] Robinhood range and the [tsymbol start date, tsymbol end date] CRSP range.

The above approach yields 8,040 tsymbols that can be mapped to a corresponding 7,969 distinct PERMNOs. Note, however, that there are two cases I need to further consider: (1) a tsymbol may correspond to more than one PERMNO over the course of the time period 05/02/2018 - 08/13/2020, and I need to identify the correct PERMNO based on the sub-period for which the tsymbol has a positive number of investors in Robinhood, i.e. the security was actively traded; and (2) a PERMNO may correspond to more than one tsymbol.

For the first case, there are six tsymbols which map to more than one different PERMNO, even after ensuring that the tsymbol start date and tsymbol end date ranges overlap with the actively traded date range of the security. These six tsymbols are CZR, GOLD, OASI, OEUR, OUSA, and SPVM. For these six tsymbols, since there is ambiguity regarding which PERMNO to use, I manually look up the company names on the Robinhood website and then use the PERMNO corresponding to the correct comnam (company name field) in CRSP. I am thus able to map these six tsymbols to their PERMNOs; the final mappings for these six securities are provided in Table IA.3.

Table IA.3
Mapping for tsymbols with multiple PERMNOs

This table provides the hand-checked PERMNO mappings for the six tsymbols that initially have more than one PERMNO match. The final matches are based on locating the closest matching company name for the security.

tsymbol	PERMNO	Company Name
CZR	13267	CAESARS ENTERTAINMENT CORP
GOLD	71298	BARRICK GOLD CORP
OASI	17861	OSI ETF TRUST
OEUR	23133	ALPS ETF TRUST (EUR)
OUSA	23134	ALPS ETF TRUST (USA)
SPVM	12876	INVESCO ETF TRUST

For the second case, I allow a PERMNO to correspond to more than one tsymbol over its lifetime since company tickers may change over time, and run all my analyses based on PERMNO rather than tsymbol. Note, however, that on any given date I mandate that a PERMNO should correspond to at most one tsymbol. This is the tsymbol that (i) has the matching PERMNO and (ii) has the appropriate [tsymbol start date, tsymbol end date] CRSP range (i.e., the tsymbol start date $\leq$ given date $\leq$ tsymbol end date condition holds).

Out of the 8,507 tsymbols from the cleaned Robintrack dataset, 8,040 (approximately 95%) are thus successfully mapped to 7,969 PERMNOs. I merge these PERMNO identifiers into the Robinhood dataset, so that now each observation contains date, tsymbol, number of Robinhood investors, first_posholding_date, last_posholding_date, namestardt, nameendt, and PERMNO (set to None if not mapped). This dataset is used in preparing the coverage time series charts displayed in Figure 1, Figure 2, and Figure 3.

B.3 Multiple Share Class NYSE-Listed Securities

My analysis identified Robintrack data discrepancies and inconsistencies in multiple share class stocks traded on the NYSE. Given that Robintrack data is becoming more widely used in academic research, I provide this appendix to highlight some of these issues that can be useful in future research.

In total, there are 48 securities in Robintrack that can be associated with a NYSE-listed multiple share class security, which I've identified through the following procedure. First, I select all security file names that contain a period in the Robintrack file name, for instance "AKO.A.csv." Next, I identify the root name of the stock, in this case "AKO." Finally, I search for this root name through all of the Robintrack filenames to identify any associated share classes of the same company. In this example, "AKO.A" and "AKO.B" are identified. The full list of all 48 share class names is shown in Table IA.4.

For illustrative purposes, the change in Robinhood investors over time in each of the 48 securities is plotted in Figure IA.1 (the first part) and Figure IA.2 (the second part). Immediately, several issues become evident. For instance, CWEN experiences wild oscillations in the number of Robinhood investors at numerous points over its history. HEI similarly has at least three large spikes in the number of Robinhood investors that almost immediately revert back, and the same pattern can be observed for GEF and other securities in the figures.

Next, I look at the underlying data for two of the problematic securities, CWEN and HEI. Note that I focus on two of the 48 securities for the sake of brevity, but the same process can be applied to catch data discrepancies in a number of the remaining 46 securities as well.

Table IA.5 demonstrates that for CWEN on 12/06/2019, Robintrack listed two different numbers of Robinhood investors in the stock simultaneously. For instance, at 12/06/2019 19:40 Robintrack has two data entries for the number of Robinhood investors in CWEN: 267 and 860. Clearly, both cannot simultaneously be accurate.

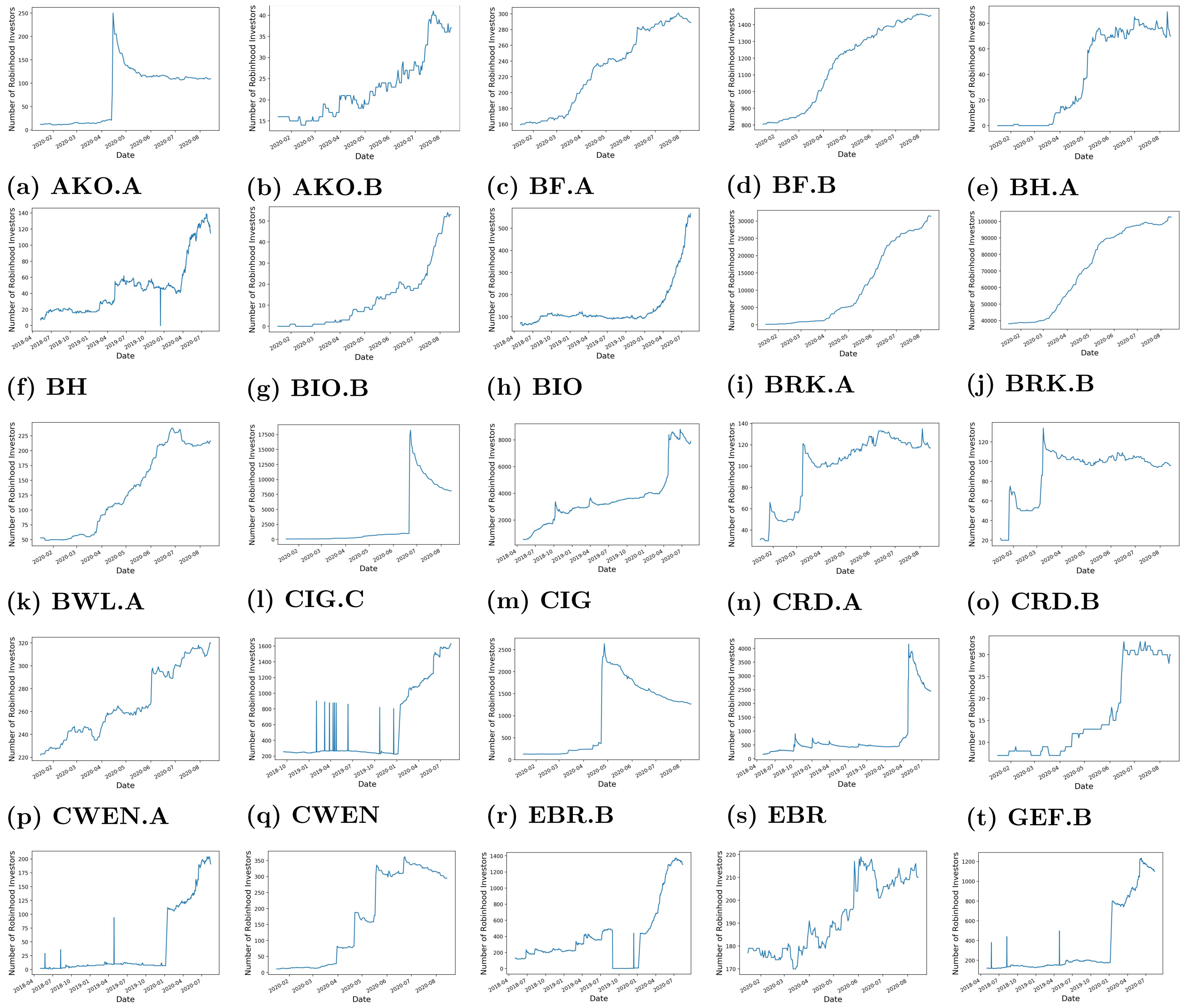
Table IA.6 provides the Robintrack data for the number of Robinhood investors in HEI and Table IA.7 provides analogous data for HEI.A. Similarly to the CWEN example, wild oscillations in the data are immediately evident. For instance, on 06/01/2020 the Robintrack data for HEI indicates ongoing swings in the number of investors between 810 and 179. Starting on 16/01/2020, the number of investors in HEI is stable at around 796 and the number of investors in HEI.A is stable at around 180, indicating that prior to 16/01/2020 it is likely that both the HEI and HEI.A numbers of Robinhood investors were being incorrectly recorded in a single HEI file. The HEI.A file also incorrectly starts only in 16/01/2020 (presumably once this issue was at least somewhat fixed in one of the Robintrack code updates) rather than earlier.

For the aforementioned reasons, I remove these multiple share class securities from the sample used in the analyses in this paper. There are only 48 such securities, which is a very small subset of the thousands of other non-multiple share class NYSE securities available in Robintrack. This appendix can be used to facilitate appropriate data cleaning of the Robintrack sample for future Robinhood research; at a minimum, the problematic dual class share stocks that have oscillating and incorrect data should be removed prior to undertaking any analysis.

Table IA.4
Multiple Share Class NYSE-Listed Securities in Robintrack

This table shows the 48 multiple share class NYSE-listed securities identified in the Robintrack data over the sample period May 2, 2018 - August 13, 2020. Company names are sourced from CRSP where available.

Ticker	Company Name
AKO.A	EMBOTELLADORA ANDINA S A
AKO.B	EMBOTELLADORA ANDINA S A
BF.A	BROWN FORMAN CORP
BF.B	BROWN FORMAN CORP
BH	BIGLARI HOLDINGS INC
BH.A	BIGLARI HOLDINGS INC
BIO	BIO RAD LABORATORIES INC
BIO.B	BIO RAD LABORATORIES INC
BRK.A	BERKSHIRE HATHAWAY INC DEL
BRK.B	BERKSHIRE HATHAWAY INC DEL
BWL.A	BOWL AMERICA INC
CIG	N/A
CIG.C	COMPANHIA ENERGETICA DE MINAS GC
CRD.A	CRAWFORD & CO
CRD.B	CRAWFORD & CO
CWEN	CLEARWAY ENERGY INC
CWEN.A	CLEARWAY ENERGY INC
EBR	CENTRAIS ELETRICAS BRASILEI
EBR.B	N/A
GEF	GREIF INC
GEF.B	GREIF INC
GTN	GRAY TELEVISION INC
GTN.A	GRAY TELEVISION INC
HEI	HEICO CORP NEW
HEI.A	HEICO CORP NEW
HVT	HAVERTY FURNITURE COS INC
HVT.A	HAVERTY FURNITURE COS INC
JW.A	WILEY JOHN & SONS INC
JW.B	WILEY JOHN & SONS INC
LEN	LENNAR CORP
LEN.B	LENNAR CORP
LGF.A	LIONS GATE ENTERTAINMENT CORP
LGF.B	LIONS GATE ENTERTAINMENT CORP
MKC	MCCORMICK & CO INC
MKC.V	MCCORMICK & CO INC
MOG.A	MOOG INC
MOG.B	MOOG INC
OIBR.C	OI S A
PBR	PETROLEO BRASILEIRO SA PETROBRAS
PBR.A	N/A
RDS.A	ROYAL DUTCH SHELL PLC
RDS.B	ROYAL DUTCH SHELL PLC
STZ	CONSTELLATION BRANDS INC
STZ.B	CONSTELLATION BRANDS INC
TAP	MOLSON COORS BREWING CO
TAP.A	MOLSON COORS BREWING CO
WSO	WATSCO INC
WSO.B	WATSCO INC



(a) **AKO.A** (b) **AKO.B** (c) **BF.A** (d) **BF.B** (e) **BH.A**

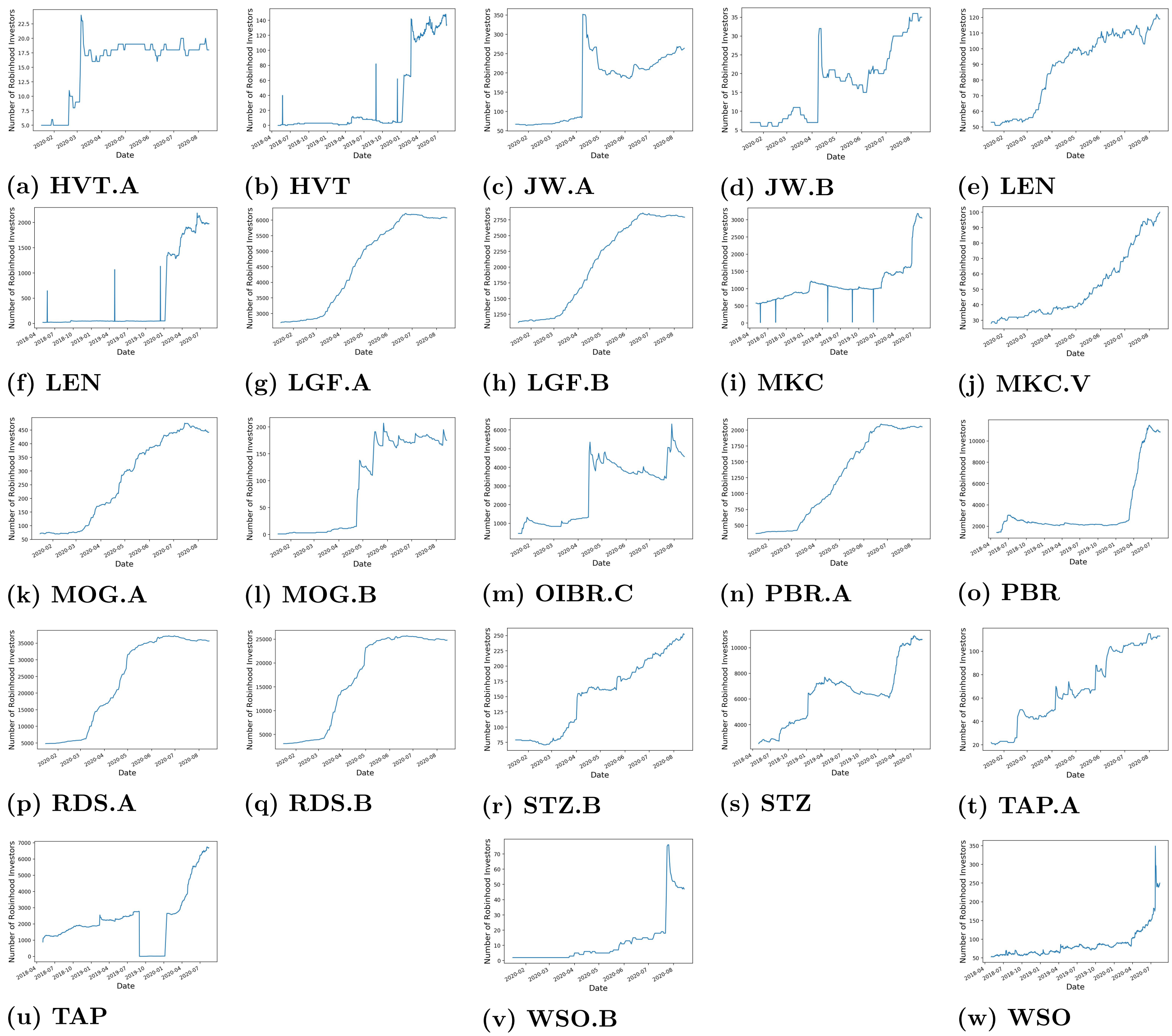
(f) **BH** (g) **BIO.B** (h) **BIO** (i) **BRK.A** (j) **BRK.B**

(k) **BWL.A** (l) **CIG.C** (m) **CIG** (n) **CRD.A** (o) **CRD.B**

(p) **CWEN.A** (q) **CWEN** (r) **EBR.B** (s) **EBR** (t) **GEF.B**

(u) **GEF** (v) **GTN.A** (w) **GTN** (x) **HEI.A** (y) **HEI**

Figure IA.1: Multiple Share Class NYSE Securities in Robintrack, Part 1 of 2
This figure shows the number of Robinhood investors per NYSE-listed multiple share class stock over time. Multiple share stocks are identified as those that contain a period (".") symbol in the filename of the security in the Robintrack file. For each sub-figure, the y axis represents the number of Robinhood investors and the x axis represents time. The Robintrack data covers the sample period May 2, 2018 - August 13, 2020 and each stock's sub-figure includes the time period for which Robintrack data is available for that stock. This figure includes the first 25 of the 48 identified NYSE multiple share class stocks in Robintrack.



(a) HVT.A (b) HVT (c) JW.A (d) JW.B (e) LEN

(f) LEN (g) LGF.A (h) LGF.B (i) MKC (j) MKC.V

(k) MOG.A (l) MOG.B (m) OIBR.C (n) PBR.A (o) PBR

(p) RDS.A (q) RDS.B (r) STZ.B (s) STZ (t) TAP.A

(u) TAP (v) WSO.B (w) WSO

Figure IA.2: Multiple Share Class NYSE Securities in Robintrack, Part 2 of 2
This figure shows the number of Robinhood investors per NYSE-listed multiple share class stock over time. Multiple share stocks are identified as those that contain a period (".") symbol in the filename of the security in the Robintrack file. For each sub-figure, the y axis represents the number of Robinhood investors and the x axis represents time. The Robintrack data covers the sample period May 2, 2018 - August 13, 2020 and each stock's sub-figure includes the time period for which Robintrack data is available for that stock. This figure includes the last 23 of the 48 identified NYSE multiple share class stocks in Robintrack.

Table IA.5
CWEN Robinhood Investors

This table shows the number of Robinhood investors in CWEN on June 12, 2019 from 17:40:00 - 20:40:00 UTC. The Robinhood investor data is from Robintrack.

Datetime (UTC)	Number of Robinhood Investors
2019-06-12 17:40:41	267
2019-06-12 17:40:43	860
2019-06-12 18:40:42	267
2019-06-12 18:40:43	860
2019-06-12 19:40:42	267
2019-06-12 19:40:43	860

Table IA.6
HEI Robinhood Investors

This table shows the number of Robinhood investors in HEI on January 06, 2020 from 06:48:00 - 08:51:00 UTC. The Robinhood investor data is from Robintrack.

Datetime (UTC)	Number of Robinhood Investors
2020-01-06 06:48:48	810
2020-01-06 06:51:38	179
2020-01-06 07:48:23	810
2020-01-06 07:50:56	179
2020-01-06 08:48:35	810

Table IA.7
HEI.A Robinhood Investors

This table shows the number of Robinhood investors in HEI.A on January 16, 2020 from 06:50:00 - 09:51:00 UTC. The Robinhood investor data is from Robintrack.

Datetime (UTC)	Number of Robinhood Investors
2020-01-16 06:50:46	180
2020-01-16 07:50:44	180
2020-01-16 08:50:42	180
2020-01-16 09:50:18	180

B.4 WallStreetBets Data

The second non-standard dataset I use in my paper is the WallStreetBets dataset. The underlying data is sourced from pushshift.io, and covers the sample period May 2018 - August 2020. This sample period coincides with the sample period for which I have Robintrack data.

In the data cleaning steps of the procedure, I first limit my observations to the WallStreetBets subreddit from all available Reddit threads. I then drop observations where both title and post content are missing. Following that, I exclude duplicated observations, identified as those observations that have the same values for the created_utc, author_created_utc, and title fields.

I next identify WallStreetBets posts that discuss a specific company; such posts are later used in the construction of the popularity and sentiment measures. The ways in which a company may be referenced in the discussion are severalfold. For instance, an individual may refer to General Electric by its ticker "GE", its common name "General Electric", or its long form name "General Electric Co." I source company tickers and long form company names from CRSP, and identify short form company names through a separate procedure. The full details of the company name matching algorithm are discussed in Appendix C.4.

C Methodology

C.1 Forward Filled Number of Investors

Robintrack suffered several data outages throughout its existence, resulting in missing data on the number of Robinhood investors.[38] For the purpose of the illustrative coverage figures, I therefore forward fill the number of Robinhood investors. The forward filled observations are not used in any of the regression analyses.

In order to construct the forward filled number of investors variable for a given stock S, I proceed as follows:

1. First, I calculate the first_date and last_date for which the number of investors holding stock S in Robintrack was positive.

2. If the first_date of positive holdings falls after May 2, 2018 (the Robintrack start date), then I backfill all the number of investors data for stock S from May 2, 2018 to first_date with 0.

3. For the number of investors data of stock S between first_date and last_date, I forward fill any missing number of investors data with the last available number of investors data point of stock S.

This procedure essentially ensures that all stocks in Robintrack have numerical data over the entire sample period May 2, 2018 - August 13, 2020, with no missing observations.

C.2 Drawdown Calculation

I calculate the maximum drawdowns for stocks over the Robinhood sample period in the following manner:

1. Step 1: I set the start date t_0 and end date t_n values. For instance, if I wish to compute the maximum drawdown over the entire sample period, I set the start date t_0 to May 2, 2018 (the first day of the sample period) and the end date t_n to August 13, 2020 (the last day of the sample period).

2. Step 2: I select the permno identifier s of the stock for which I wish to compute the drawdown.

3. Step 3: For each daily observation $t_i \in \{t_0, t_n\}$ between the start date and end date, I compute $ret^c_{t_i,s}$, the cumulative return of the stock s from the start date t_0 up through t_i. The daily returns data is from CRSP.

4. Step 4: Next, for each date t_i I compute the maximum cumulative return reached by stock s at some point on or before t_i. This is going to be equal to the highest cumulative return reached up to that date, and for the purposes of calculating the max drawdown it would represent a possible "worst point" for entering into a position in the given stock (a point of its highest valuation). More formally, $ret^{c,max}_{t_i,s} = \max_{t \in \{t_0,..,t_i\}} ret^c_{t,s}$.

5. Step 5: I then compute the drawdown value at each date t_i if the investor had invested at the worst possible previous point. To do so, I divide 1 plus the current cumulative return by 1 plus the highest cumulative return reached up to that date (calculated in Step 4 above) and subtract

[38]Specifically, Robintrack failed to gather the number of investors data over the periods January 25-29, 2019 and January 7-15, 2020.

1, so that $dd_{t_i,s} = (1 + ret^c_{t_i,s})/(1 + ret^{c,max}_{t_i,s}) - 1$. Essentially this computes the drawdown if the investor had invested at the highest previous point and sold their investment on the current date t_i.

6. Step 6: Finally, armed with the daily drawdown values, to calculate the maximum possible drawdown as of each date (where the investor invested not only at the highest possible previous point but also sold at the lowest possible previous point) I compute $\min_{t \in \{t_0,..,t_i\}} dd_{t,s}$. Note that the minimum operator is used to find the maximum drawdown since drawdown values are negative values, and I am looking for the drawdown of the largest magnitude.

7. Step 7: The maximum drawdown over the full sample period from start date t_0 to end date t_n is then the minimum (i.e., most negative) of all of the daily maximum possible drawdown values.

For completeness, note that there are multiple ways of measuring strategy drawdowns in the applied fund industry literature. Here I have outlined one such method, which I use in my drawdown analysis.

C.3 Number of Robinhood Investors Standardization Methods

In addition to using the log winsorized value of the number of Robinhood investors used in the majority of this paper's analysis, I also consider a number of alternative approaches. Alternative methods of standardizing the number of investors in each security include: (1) demeaning the number of investors in stock i at time t against the average number of investors in all other Robinhood securities at time t; (2) calculating the proportion that the number of investors in stock i at time t constitute out of the average number of Robinhood investors in a Dow Jones stock at time t; (3) calculating the proportion that the number of Robinhood investors in stock i at time t constitute out of the number of Robinhood investors in the S&P 500 ETF at time t; and (4) using the percentile of the number of Robinhood investors in stock i at time t relative to the number of Robinhood investors in all other stocks at time t.

I find that of the methods considered, the log winsorized number of investors and the percentile of investors yield an optimal balance between retaining important information and reducing the impact of missing Robintrack data and outliers. I perform the analyses in this paper using percentile as a robustness check, and find similar results.

C.4 Company Name Matching Algorithm

In order to measure the frequency of discussions about a particular company on the WallStreetBets investing subreddit, I search post titles and content for the occurrence of either (a) the company's ticker preceded by the $ sign or (b) the company's name. If either the company's ticker or the company's name are mentioned in the post's title or content, I mark the post as one that talks about the given company.

The mapping for company tickers is relatively straightforward, since it is a common practice on the WallStreetBets subreddit to precede ticker mentions with the dollar sign to avoid confusion, which in turn makes tickers more easily parsed. However, some difficulties arise when searching for a company name for the simple reason that many companies' official incorporated names are not the same as those commonly used by individuals in more informal discussions. For example, the companies with CRSP company names "APPLE INC", "MICROSOFT CORP", "ALPHABET INC", "A T & T INC", and "SALESFORCE COM INC" are more commonly known as "Apple", "Microsoft", "Google", "AT&T", and "Salesforce" respectively.

In order to map CRSP company names to their more commonly used company name variations, I proceed as follows:

1. For a given company name in CRSP, I first remove all instances of " INC", " LTD", " MD", " GROUP", " HOLDINGS", " HLDGS", " INTL", " PLC", and " & CO" in the name.

2. In the second step, I remove all instances of " NEW", " COM", " CO", " CORP", " COMPANY", " INTERNATIONAL", " COMMUNICATIONS", " TECHNOLOGIES", " HOLDING", " SYSTEM", " GRP", " SVCS", " SOLS", " OH", " GP", " USA", " SA", " SF", and " DEL" where those phrases occur at the end of the company name. In order to achieve this, I loop over the company names until no such phrases at the end of the company names remain (note that this is key in order to properly adjust company names that include several of these phrases simultaneously, such as "COMCAST CORP NEW"). In addition, it is imperative not to remove the phrases if they occur in the middle of the company name since they may then be part of a broader, more easily recognized name.

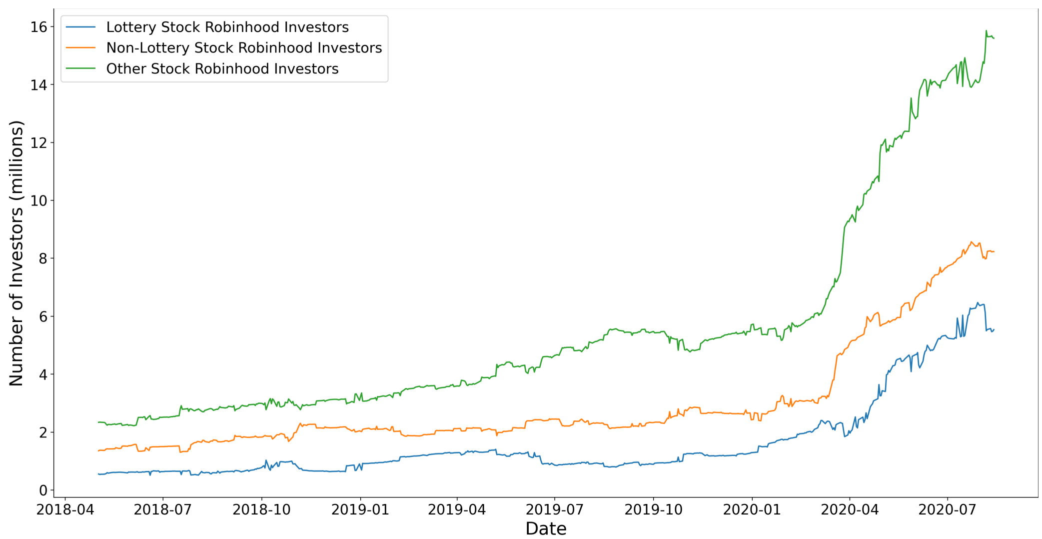
3. I proceed by combining individual letter segments in the company name together to form words. For instance, at this stage "A T & T" would be transformed into "AT&T."

4. I spell out common CRSP abbreviations fully, such as converting instances of "SOFTWR" to "software", "SOLUTNS" to "SOLUTIONS", and "SECUR" to "SECURITY".

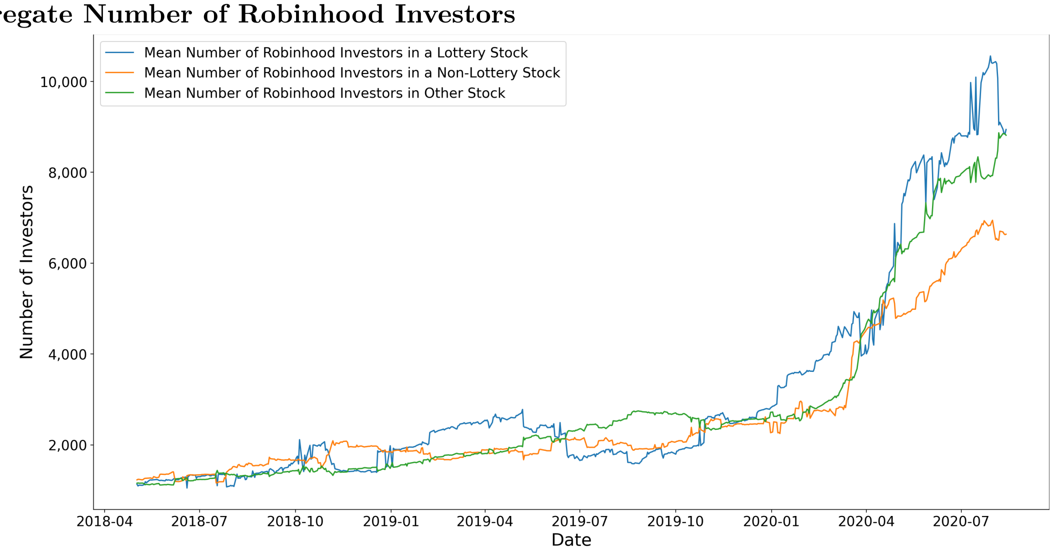
5. I next handle one-off special cases, such as transforming "MORGAN STANLEY DEAN WITTER" to "MORGAN STANLEY" and mapping "ALPHABET" to "GOOGLE".

6. Finally, I employ title style capitalization, so that all short company names follow a capitalization structure where the first letter of every key word is capitalized. At this point, "MORGAN STANLEY" from the previous example would become "Morgan Stanley".

D Robustness Tables and Figures



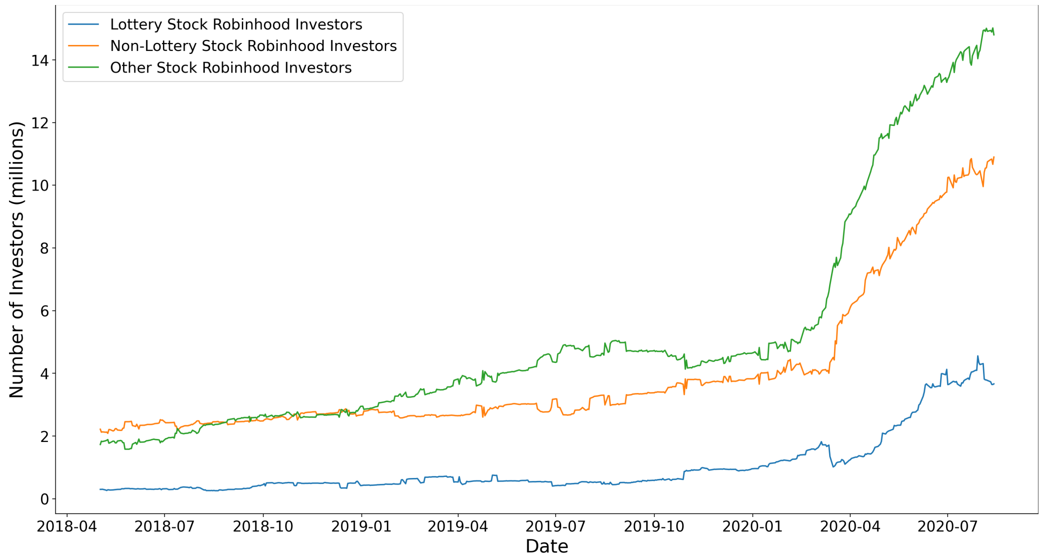
(a) Aggregate Number of Robinhood Investors



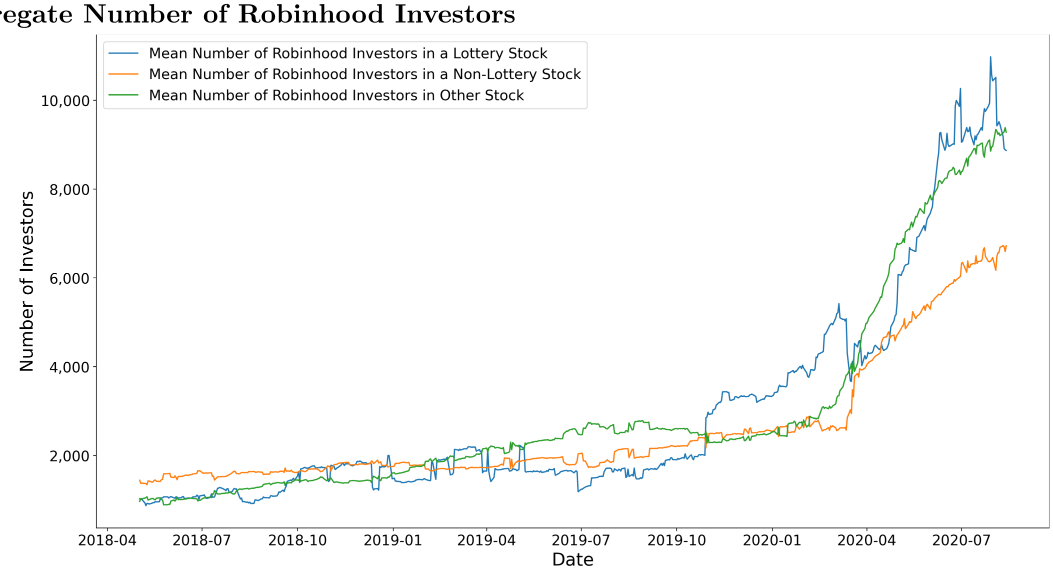
(b) Mean Robinhood Investors per Stock

Figure IA.3: Robinhood Investors in Lottery Stocks, Non-Lottery Stocks, and Other Stocks ($k = 40$)

This chart plots a time series of the total number of Robinhood investors (Panel (a)) and mean number of Robinhood investors per stock (Panel (b)) in lottery stocks, non-lottery stocks, and other stocks over time. Lottery stocks are defined as those stocks whose price falls into the lowest 40th percentile, whose idiosyncratic volatility falls into the highest 40th percentile, and whose idiosyncratic skewness falls into the highest 40th percentile. Non-lottery stocks are defined as those stocks whose price does not fall into the lowest 40th percentile, whose idiosyncratic volatility does not fall into the highest 40th percentile, and whose idiosyncratic skewness does not fall into the highest 40th percentile. Other stocks include all remaining stocks.



(a) Aggregate Number of Robinhood Investors



(b) Mean Robinhood Investors per Stock

Figure IA.4: Robinhood Investors in Lottery Stocks, Non-Lottery Stocks, and Other Stocks ($k = 30$)

This chart plots a time series of the total number of Robinhood investors (Panel (a)) and mean number of Robinhood investors per stock (Panel (b)) in lottery stocks, non-lottery stocks, and other stocks over time. Lottery stocks are defined as those stocks whose price falls into the lowest 30th percentile, whose idiosyncratic volatility falls into the highest 30th percentile, and whose idiosyncratic skewness falls into the highest 30th percentile. Non-lottery stocks are defined as those stocks whose price does not fall into the lowest 30th percentile, whose idiosyncratic volatility does not fall into the highest 30th percentile, and whose idiosyncratic skewness does not fall into the highest 30th percentile. Other stocks include all remaining stocks.

Table IA.8
Representative Portfolio Sector Weights - Fama French 49 Industries

This table shows the Fama French 49 Industry Classification weights for the market portfolio, Robinhood (dollar method) portfolio, and Robinhood (share method) portfolio as of August 13, 2020 (the Robintrack sample period end date). Historical SIC information is from CRSP. Fama French 49 Industry Classifications come from Kenneth R. French's data library.

Fama French Industry	Market Portfolio Weight	Robinhood (Dollar Method) Portfolio Weight	Robinhood (Share Method) Portfolio Weight
Agriculture	0.08%	0.02%	0.00%
Aircraft	1.32%	2.13%	2.00%
Almost Nothing	0.30%	0.18%	0.12%
Apparel	0.71%	0.88%	0.41%
Automobiles and Trucks	0.56%	4.83%	0.35%
Banking	4.66%	3.92%	1.04%
Beer and Liquor	0.27%	0.03%	0.07%
Business Services	26.11%	18.23%	60.08%
Business Supplies	0.38%	0.10%	0.04%
Candy and Soda	0.98%	0.21%	0.17%
Chemicals	1.04%	0.36%	0.16%
Coal	0.01%	0.03%	0.00%
Communication	3.61%	3.44%	0.93%
Computers	7.41%	3.92%	10.62%
Construction	0.45%	0.16%	0.13%
Construction Materials	0.64%	0.12%	0.05%
Consumer Goods	2.23%	1.20%	0.35%
Defense	0.38%	0.24%	0.28%
Electrical Equipment	0.30%	2.60%	0.25%
Electronic Equipment	5.20%	4.15%	3.80%
Entertainment	1.56%	6.17%	6.32%
Fabricated Products	0.02%	0.02%	0.00%
Food Products	1.11%	0.34%	0.17%
Healthcare	0.58%	0.45%	0.28%
Insurance	5.18%	0.44%	0.34%
Machinery	2.19%	4.20%	0.45%
Measuring and Control Equipment	2.06%	1.46%	0.34%
Medical Equipment	3.78%	1.24%	1.04%
Non-Metallic and Industrial Metal Mining	0.30%	0.54%	0.03%
Personal Services	0.17%	0.81%	0.08%
Petroleum and Natural Gas	2.32%	6.54%	0.74%
Pharmaceutical Products	5.86%	6.60%	1.62%
Precious Metals	0.18%	0.31%	0.03%
Printing and Publishing	0.06%	0.33%	0.03%
Real Estate	0.23%	0.17%	0.05%
Recreation	0.11%	0.33%	0.08%
Restaurants, Hotels, Motels	1.64%	2.98%	1.63%
Retail	5.61%	6.00%	2.72%
Rubber and Plastic Products	0.21%	0.11%	0.02%
Shipbuilding, Railroad Equipment	0.07%	0.02%	0.01%
Shipping Containers	0.14%	0.02%	0.01%
Steel Works Etc	0.12%	0.23%	0.02%
Textiles	0.05%	0.06%	0.01%
Tobacco Products	0.65%	0.19%	0.07%
Trading	2.90%	0.95%	0.48%
Transportation	1.97%	10.60%	2.00%
Utilities	3.23%	1.25%	0.35%
Wholesale	1.06%	0.89%	0.23%

Table IA.9

Lottery Stock, Value, and Size Investment Preferences - 2018

This table reports the panel regression estimates with time fixed effects in columns (1), (2), (3), and (4) and the Fama and MacBeth (1973) cross-sectional regression estimates in columns (5) and (6) for the regressions of the number of Robinhood investors on lottery stock characteristics, size, and value in the sample period limited to the calendar year 2018. The dependent variable is computed as the log winsorized number of Robinhood investors in stock i at time t. All independent variables are measured at the end of the previous day. The control variables include momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, dividends per share, an indicator variable for a dividend-paying firm, firm age, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. The independent variables of interest are standardized. Regressions include time fixed effects and standard errors are clustered by time and security.

	Number of Robinhood Investors					
	(1)	(2)	(3)	(4)	(5)	(6)
Stock price	4.23			9.95***	4.23***	9.50***
	(3.31)			(2.47)	(0.09)	(0.04)
Idiosyncratic volatility	0.52***			0.29***	0.52***	0.29***
	(0.09)			(0.09)	(0.01)	(0.00)
Idiosyncratic skewness	-0.08***			-0.01	-0.09***	-0.02***
	(0.03)			(0.02)	(0.01)	(0.00)
ln(Size)		0.59***		0.12		0.09***
		(0.04)		(0.10)		(0.01)
Book-to-market ratio			-0.39***	-0.19***		-0.19***
			(0.06)	(0.03)		(0.00)
Controls	No	No	No	Yes	No	Yes
R-squared	0.05	0.10	0.03	0.66	0.05	0.67
Time fixed effects	Yes	Yes	Yes	Yes	-	-
Number of observations	456,472	498,995	470,698	433,290	456,472	433,290

Table IA.10
Lottery Stock, Value, and Size Investment Preferences - 2019

This table reports the panel regression estimates with time fixed effects in columns (1), (2), (3), and (4) and the Fama and MacBeth (1973) cross-sectional regression estimates in columns (5) and (6) for the regressions of the number of Robinhood investors on lottery stock characteristics, size, and value in the sample period limited to the calendar year 2019. The dependent variable is computed as the log winsorized number of Robinhood investors in stock i at time t. All independent variables are measured at the end of the previous day. The control variables include momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, dividends per share, an indicator variable for a dividend-paying firm, firm age, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. The independent variables of interest are standardized. Regressions include time fixed effects and standard errors are clustered by time and security.

	Number of Robinhood Investors					
	(1)	(2)	(3)	(4)	(5)	(6)
Stock price	3.45			6.74***	3.53***	6.48***
	(2.92)			(1.71)	(0.06)	(0.07)
Idiosyncratic volatility	0.57***			0.20***	0.59***	0.25***
	(0.09)			(0.08)	(0.01)	(0.01)
Idiosyncratic skewness	-0.16***			-0.06***	-0.16***	-0.06***
	(0.02)			(0.02)	(0.00)	(0.00)
ln(Size)		0.52***		0.12		0.10***
		(0.03)		(0.08)		(0.01)
Book-to-market ratio			-0.28***	-0.15***		-0.15***
			(0.04)	(0.03)		(0.00)
Controls	No	No	No	Yes	No	Yes
R-squared	0.06	0.08	0.02	0.63	0.06	0.64
Time fixed effects	Yes	Yes	Yes	Yes	-	-
Number of observations	724,498	803,522	749,132	687,530	724,498	687,530

Table IA.11

Lottery Stock, Value, and Size Investment Preferences - 2020

This table reports the panel regression estimates with time fixed effects in columns (1), (2), (3), and (4) and the Fama and MacBeth (1973) cross-sectional regression estimates in columns (5) and (6) for the regressions of the number of Robinhood investors on lottery stock characteristics, size, and value in the sample period limited to the calendar year 2020. The dependent variable is computed as the log winsorized number of Robinhood investors in stock i at time t. All independent variables are measured at the end of the previous day. The control variables include momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, dividends per share, an indicator variable for a dividend-paying firm, firm age, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. The independent variables of interest are standardized. Regressions include time fixed effects and standard errors are clustered by time and security.

	Number of Robinhood Investors					
	(1)	(2)	(3)	(4)	(5)	(6)
Stock price	6.86**			8.84***	6.99***	8.32***
	(2.84)			(1.62)	(0.28)	(0.10)
Idiosyncratic volatility	0.54***			0.22***	0.55***	0.23***
	(0.04)			(0.04)	(0.02)	(0.00)
Idiosyncratic skewness	-0.12***			-0.06***	-0.11***	-0.02***
	(0.03)			(0.02)	(0.01)	(0.00)
ln(Size)		0.49***		0.16**		0.12***
		(0.03)		(0.07)		(0.01)
Book-to-market ratio			-0.18***	-0.09***		-0.08***
			(0.03)	(0.02)		(0.00)
Controls	No	No	No	Yes	No	Yes
R-squared	0.12	0.09	0.04	0.69	0.09	0.70
Time fixed effects	Yes	Yes	Yes	Yes	-	-
Number of observations	474,088	521,047	482,556	444,676	474,088	444,676

Table IA.12

Lottery Stock Investment Preferences - Percentile Method

This table reports the panel regression estimates with time fixed effects in columns (1), (2), (3), and (4) and the Fama and MacBeth (1973) cross-sectional regression estimates in columns (5) and (6) for the regressions of the percentile of Robinhood investors on lottery stock characteristics, size, and value in the sample period limited to the calendar year 2020. The dependent variable is computed as the percentile of the number of Robinhood investors in stock i at time t. All independent variables are measured at the end of the previous day. The control variables include momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, dividends per share, an indicator variable for a dividend-paying firm, firm age, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. The independent variables of interest are standardized. Regressions include time fixed effects and standard errors are clustered by time and security.

	Percentile of Robinhood Investors			
	(1)	(2)	(3)	(4)
Stock price	0.54	1.19***	0.55***	1.14***
	(0.39)	(0.25)	(0.02)	(0.01)
Idiosyncratic volatility	0.09***	0.03***	0.10***	0.04***
	(0.01)	(0.01)	(0.00)	(0.00)
Idiosyncratic skewness	-0.02***	-0.01***	-0.02***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Controls	No	Yes	No	Yes
R-squared	0.08	0.64	0.08	0.65
Time fixed effects	Yes	Yes	-	-
Number of observations	1,655,058	1,565,496	1,655,058	1,565,496

Table IA.13
Lottery Stock Investment Preferences - Percentile Method, by Year

This table reports the panel regression estimates with time fixed effects in columns (1), (2), (3), and (4) and the Fama and MacBeth (1973) cross-sectional regression estimates in columns (5) and (6) for the regressions of the percentile of Robinhood investors on lottery stock characteristics, size, and value. Panel A, Panel B, and Panel C provide estimates for the calendar years 2018, 2019, and 2020 respectively. The dependent variable is computed as the percentile of the number of Robinhood investors in stock i at time t. All independent variables are measured at the end of the previous day. The control variables include momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, dividends per share, an indicator variable for a dividend-paying firm, firm age, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. The independent variables of interest are standardized. Regressions include time fixed effects and standard errors are clustered by time and security.

Panel A: 2018

	Percentile of Robinhood Investors			
	(1)	(2)	(3)	(4)
Stock price	0.50	1.52***	0.50***	1.45***
	(0.44)	(0.34)	(0.01)	(0.01)
Idiosyncratic volatility	0.09***	0.04***	0.09***	0.04***
	(0.02)	(0.01)	(0.00)	(0.00)
Idiosyncratic skewness	-0.02***	-0.00	-0.02***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Controls	No	Yes	No	Yes
R-squared	0.06	0.65	0.06	0.66
Time fixed effects	Yes	Yes	-	-
Number of observations	456,472	433,290	456,472	433,290

Panel B: 2019

	Percentile of Robinhood Investors			
	(1)	(2)	(3)	(4)
Stock price	0.39	0.97***	0.41***	0.94***
	(0.39)	(0.23)	(0.01)	(0.01)
Idiosyncratic volatility	0.10***	0.03***	0.10***	0.04***
	(0.02)	(0.01)	(0.00)	(0.00)
Idiosyncratic skewness	-0.03***	-0.01***	-0.03***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Controls	No	Yes	No	Yes
R-squared	0.08	0.62	0.08	0.63
Time fixed effects	Yes	Yes	-	-
Number of observations	724,498	687,530	724,498	687,530

Panel C: 2020

	Percentile of Robinhood Investors			
	(1)	(2)	(3)	(4)
Stock price	0.84**	1.21***	0.85***	1.13***
	(0.37)	(0.24)	(0.04)	(0.01)
Idiosyncratic volatility	0.08***	0.03***	0.09***	0.03***
	(0.01)	(0.01)	(0.00)	(0.00)
Idiosyncratic skewness	-0.02***	-0.01**	-0.01***	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Controls	No	Yes	No	Yes
R-squared	0.10	0.67	0.10	0.68
Time fixed effects	Yes	Yes	-	-
Number of observations	474,088	444,676	474,088	444,676

Table IA.14

Lottery Stock, Value, and Size Investment Preferences - Total Measures

This table reports the panel regression estimates with time fixed effects in columns (1), (2), (3), and (4) and the Fama and MacBeth (1973) cross-sectional regression estimates in columns (5) and (6) for the regressions of the number of Robinhood investors on lottery stock characteristics, size, and value in the sample period limited to the calendar year 2018. The dependent variable is computed as the log winsorized number of Robinhood investors in stock i at time t. All independent variables are measured at the end of the previous day. The control variables include momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, dividends per share, an indicator variable for a dividend-paying firm, firm age, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. The independent variables of interest are standardized. Regressions include time fixed effects and standard errors are clustered by time and security.

	Number of Robinhood Investors					
	(1)	(2)	(3)	(4)	(5)	(6)
Stock price	4.92			8.30^{***}	4.99^{***}	7.86^{***}
	(3.03)			(1.81)	(0.14)	(0.07)
Total volatility	0.62^{***}			0.23^{***}	0.64^{***}	0.26^{***}
	(0.06)			(0.04)	(0.01)	(0.00)
Total skewness	-0.16^{***}			-0.05^{***}	-0.16^{***}	-0.03^{***}
	(0.02)			(0.01)	(0.01)	(0.00)
ln(Size)		0.53^{***}		0.12^{*}		0.10^{***}
		(0.03)		(0.07)		(0.00)
Book-to-market ratio			-0.26^{***}	-0.14^{***}		-0.14^{***}
			(0.03)	(0.02)		(0.00)
Controls	No	No	No	Yes	No	Yes
R-squared	0.12	0.12	0.07	0.67	0.08	0.67
Time fixed effects	Yes	Yes	Yes	Yes	-	-
Number of observations	1,766,919	1,823,564	1,702,386	1,565,496	1,766,919	1,565,496

Table IA.15

Lottery Stock, Value, and Size Investment Preferences with Industry Fixed Effects

This table reports the panel regression estimates with time fixed effects and industry fixed effects in columns (1), (2), (3), and (4) and the Fama and MacBeth (1973) cross-sectional regression estimates in columns (5) and (6) for the regressions of the number of Robinhood investors on lottery stock characteristics, size, and value. The dependent variable is computed as the log winsorized number of Robinhood investors in stock i at time t. All independent variables are measured at the end of the previous day. The control variables include momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, dividends per share, an indicator variable for a dividend-paying firm, firm age, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. Industry is measured using the NAICS industry classification system. The independent variables of interest are standardized. Regressions include time fixed effects and industry fixed effects, and standard errors are clustered by time and security.

	Number of Robinhood Investors					
	(1)	(2)	(3)	(4)	(5)	(6)
Stock price	3.87*			7.63***	4.65***	7.86***
	(2.21)			(1.71)	(0.16)	(0.07)
Idiosyncratic volatility	0.35***			0.22***	0.56***	0.25***
	(0.04)			(0.04)	(0.01)	(0.00)
Idiosyncratic skewness	-0.09***			-0.06***	-0.12***	-0.04***
	(0.01)			(0.01)	(0.00)	(0.00)
ln(Size)		0.53***		0.20***		0.10***
		(0.03)		(0.07)		(0.00)
Book-to-market ratio			-0.16***	-0.09***		-0.14***
			(0.03)	(0.02)		(0.00)
Controls	No	No	No	Yes	No	Yes
R-squared	0.26	0.31	0.24	0.69	0.07	0.67
Time fixed effects	Yes	Yes	Yes	Yes	-	-
Industry fixed effects	Yes	Yes	Yes	Yes	-	-
Number of observations	1,642,492	1,703,155	1,697,161	1,561,554	1,655,058	1,565,496

Table IA.16

Lottery Stock, Value, and Size Investment Preferences with Industry Fixed Effects

This table reports the panel regression estimates with time fixed effects and industry fixed effects in columns (1), (2), (3), and (4) and the Fama and MacBeth (1973) cross-sectional regression estimates in columns (5) and (6) for the regressions of the number of Robinhood investors on lottery stock characteristics, size, and value. The dependent variable is computed as the log winsorized number of Robinhood investors in stock i at time t. All independent variables are measured at the end of the previous day. The control variables include momentum, gross profitability, investment, systematic volatility, systematic skewness, total kurtosis, dividends per share, an indicator variable for a dividend-paying firm, firm age, share and dollar volume traded, Amihud illiquidity, and bid-ask spread. Industry is measured using the NAICS industry classification system. The independent variables of interest are standardized. Regressions include time fixed effects and industry fixed effects, and standard errors are clustered by time and security.

	Number of Robinhood Investors					
	(1)	(2)	(3)	(4)	(5)	(6)
Stock Price	3.87*			7.63***	4.65***	7.86***
	(2.21)			(1.71)	(0.16)	(0.07)
Idiosyncratic Volatility	0.35***			0.22***	0.56***	0.25***
	(0.04)			(0.04)	(0.01)	(0.00)
Idiosyncratic Skewness	-0.09***			-0.06***	-0.12***	-0.04***
	(0.01)			(0.01)	(0.00)	(0.00)
ln(Size)		0.53***		0.20***		0.10***
		(0.03)		(0.07)		(0.00)
Book-to-market Ratio			-0.16***	-0.09***		-0.14***
			(0.03)	(0.02)		(0.00)
Controls	No	No	No	Yes	No	Yes
R-squared	0.26	0.31	0.24	0.69	0.07	0.67
Time-Fixed Effects	Yes	Yes	Yes	Yes	-	-
Industry-Fixed Effects	Yes	Yes	Yes	Yes	-	-
Number of observations	1,642,492	1,703,155	1,697,161	1,561,554	1,655,058	1,565,496

Table IA.17
Three-Day Change in the Number of Robinhood Investors by Cumulative Abnormal Return (CAR) Quintile, CAPM Alphas

This table shows the average three-day percent change in the number of Robinhood investors for stocks in a given cumulative abnormal return (CAR) quintile. CAR is computed by compounding the contemporaneous three-day CAPM alpha. Quintile 1 represents the lowest CAR quintile and Quintile 5 represents the highest CAR quintile. Small Cap includes all observations for stocks with a market cap less than or equal to $2bn, Mid Cap includes all observations for stocks with a market cap between $2bn and $10bn, and Large Cap includes all observations for stocks with a market cap over $10bn. The Overall column includes observations across all market caps, and the Observation Count column reports the number of observations used to compute the quintile averages in the Overall column.

CAR Quintile	Small Cap	Mid Cap	Large Cap	Overall	Observation Count
Quintile 1	1.5%	2.3%	3.0%	1.8%	368,923
Quintile 2	0.3%	0.3%	0.6%	0.3%	324,468
Quintile 3	0.2%	0.2%	0.3%	0.2%	304,214
Quintile 4	0.3%	0.2%	0.3%	0.2%	320,120
Quintile 5	2.7%	1.4%	1.1%	2.2%	363,699

E Extensions

E.1 Investment in Other Security Types

In this appendix, I provide descriptive statistics for non-U.S. equity direct investment securities on the Robinhood platform. Table IA.18 shows the top 10 holdings of Robinhood investors in foreign stocks, ETFs, U.S. mutual funds, ADRs, and REITs. The maximum drawdown is computed as outlined in Appendix C.2. The table is indicative of the fact that for other security types, Robinhood investors also appear to have traded in adherence with the buy-the-dip effect. This is evidenced by the consistently large magnitudes of the maximum drawdowns for the top holdings, particularly for foreign stocks.

Table IA.18
Top Holdings of Robinhood Investors by Security Type

This table shows the top 10 holdings among Robinhood investors for different security types as of August 13, 2020, the Robintrack sample end date. Maximum drawdown is computed as the maximum drawdown an investor may have experienced in the security if they had entered and exited positions at the most inopportune dates in the May 2, 2018 - August 13, 2020 sample period. Panel A displays the top holdings for foreign stocks, Panel B displays the top holdings for ETFs, Panel C displays the top holdings for U.S. mutual funds, Panel D displays the top holdings for ADRs, and Panel E displays the top holdings for REITs.

Panel A - Foreign Stocks

Company Name	Ticker	Number of Investors	Maximum Drawdown
Aurora Cannabis	ACB	430,719	-95.15%
Norwegian Cruise Line	NCLH	353,238	-86.97%
Hexo	HEXO	248,822	-95.61%
Canopy Growth	CGC	233,413	-82.90%
Royal Caribbean Cruises	RCL	221,171	-83.27%
Cronos	CRON	217,879	-80.93%
Aphria	APHA	210,874	-83.89%
Organigram	OGI	178,136	-85.10%
Top Ships	TOPS	152,708	-99.87%
Pacific Drilling Sa	PACD	74,889	-97.72%

Panel B - ETFs

Company Name	Ticker	Number of Investors	Maximum Drawdown
Proshares Trust Ii	UCO	137,267	-98.75%
Vanguard Index Funds	VOO	136,380	-34.01%
Spdr S&P 500 Etf Trust	SPY	118,656	-33.70%
Invesco Etf Trust Ii	SPHD	90,533	-41.36%
Vanguard Index Funds	VTI	69,349	-35.00%
Invesco Qqq Trust	QQQ	49,257	-28.56%
Vanguard Whitehall Funds	VYM	48,798	-35.22%
Etf Managers Trust	MJ	45,695	-77.35%
Direxion Shares Etf Trust	JNUG	41,928	-96.69%
Spdr Gold Trust	GLD	40,522	-12.53%

Panel C - U.S. Mutual Funds

Company Name	Ticker	Number of Investors	Maximum Drawdown
United States Oil Fund Lp	USO	144,990	-86.75%
Prospect Capital	PSEC	76,497	-41.16%
Main Street Capital	MAIN	18,057	-64.50%
Oxford Lane Capital	OXLC	16,980	-74.42%
Gladstone Investment	GAIN	13,029	-56.30%
Gladstone Capital	GLAD	6,941	-58.49%
Oxford Square Capital	OXSQ	5,944	-67.03%
Capitala Finance	CPTA	5,737	-77.14%
Duff & Phelps Sel Mm Enr Fd	DSE	5,580	-93.91%
Cornerstone Strategic Val Fd	CLM	5,484	-45.06%

Panel D - ADRs

Company Name	Ticker	Number of Investors	Maximum Drawdown
Nio	NIO	268,489	-88.62%
Alibaba	BABA	255,282	-38.06%
Nokia	NOK	130,837	-62.57%
Sony	SNE	110,273	-30.70%
Bp	BP	74,800	-62.48%
Astrazeneca	AZN	74,223	-24.83%
Kitov Pharma	KTOV	49,217	-92.41%
Jd	JD	41,558	-55.96%
Bilibili	BILI	40,092	-51.76%
Ericsson	ERIC	38,876	-40.62%

Panel E - REITs

Company Name	Ticker	Number of Investors	Maximum Drawdown
Mfa Financial	MFA	150,539	-95.52%
Invesco Mortgage Capital	IVR	137,609	-88.18%
New Residential Investment	NRZ	112,443	-81.11%
New York Mortgage Trust	NYMT	76,729	-84.01%
Realtyome	O	63,815	-48.28%
Two Harbors Investment	TWO	53,283	-84.71%
Washington Prime	WPG	49,943	-89.68%
Simon Property	SPG	35,765	-75.35%
Apple Hospitality Reit	APLE	33,027	-71.13%
Redwood Trust	RWT	32,945	-85.40%

F Proofs

F.1 Mean-Variance Portfolio Rebalancing

Consider the case of a mean-variance efficient portfolio composed of three assets. Let the mean return of the first asset be μ_1, and the mean return of the second asset be μ_2. Let the variance-covariance matrix be $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}$. Denote the weights of the three assets in the portfolio as x_1, x_2, and x_3, and note that $x_1 + x_2 + x_3 = 1$ (the sum of the portfolio weights is equal to one) and $x_1 \geq 0, x_2 \geq 0, x_3 \geq 0$ (no shorting).

We can write $x_3 = 1 - x_1 - x_2$. For a given expected return E and variance V, we then have:

$$\begin{aligned} E &= x_1\mu_1 + x_2\mu_2 + x_3\mu_3 \\ &= x_1\mu_1 + x_2\mu_2 + (1 - x_1 - x_2)\mu_3 \\ &= \mu_3 + x_1(\mu_1 - \mu_3) + x_2(\mu_2 - \mu_3) \end{aligned} \tag{19}$$

Similarly, substituting the expression for x_3, the variance of the portfolio can be written as:

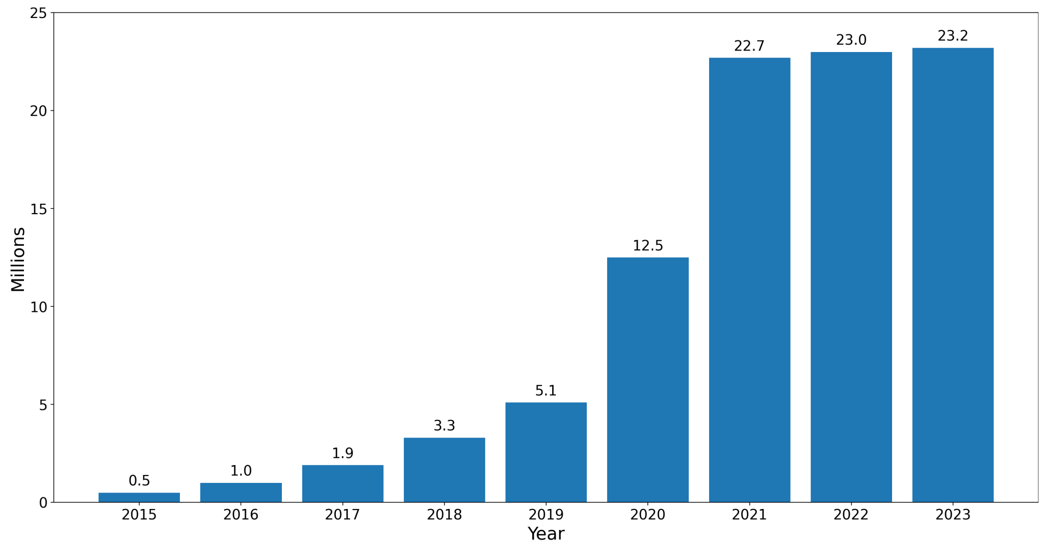
$$\begin{aligned} V &= x_1^2\sigma_1^2 + x_2^2\sigma_2^2 + x_3^2\sigma_3^2 + 2x_1x_2\sigma_{12} + 2x_1x_3\sigma_{13} + 2x_2x_3\sigma_{23} \\ &= x_1^2\sigma_1^2 + x_2^2\sigma_2^2 + (1 - x_1 - x_2)^2\sigma_3^2 + 2x_1x_2\sigma_{12} + 2x_1(1 - x_1 - x_2)\sigma_{13} + 2x_2(1 - x_1 - x_2)\sigma_{23} \\ &= x_1^2(\sigma_{11} + \sigma_{33} - 2\sigma_{13}) + x_2^2(\sigma_{22} + \sigma_{33} - 2\sigma_{23}) + 2x_1x_2(\sigma_{33} + \sigma_{12} - \sigma_{13} - \sigma_{23}) \\ &+ 2x_1(\sigma_{13} - \sigma_{33}) + 2x_2(\sigma_{23} - \sigma_{33}) + \sigma_{33} \end{aligned} \tag{20}$$

Given a target expected return e, equation (19) can be used to solve for x_2 in terms of x_1 so that $x_2 = \frac{e - \mu_3 - x_1(\mu_1 - \mu_3)}{\mu_2 - \mu_3}$. Substituting that expression into the equation for variance (20) yields a single solution for x_1 that minimizes the variance of the portfolio. This solution for x_1, x_2 and x_3 represents the optimal portfolio weights for the minimum variance portfolio for a given expected return e.

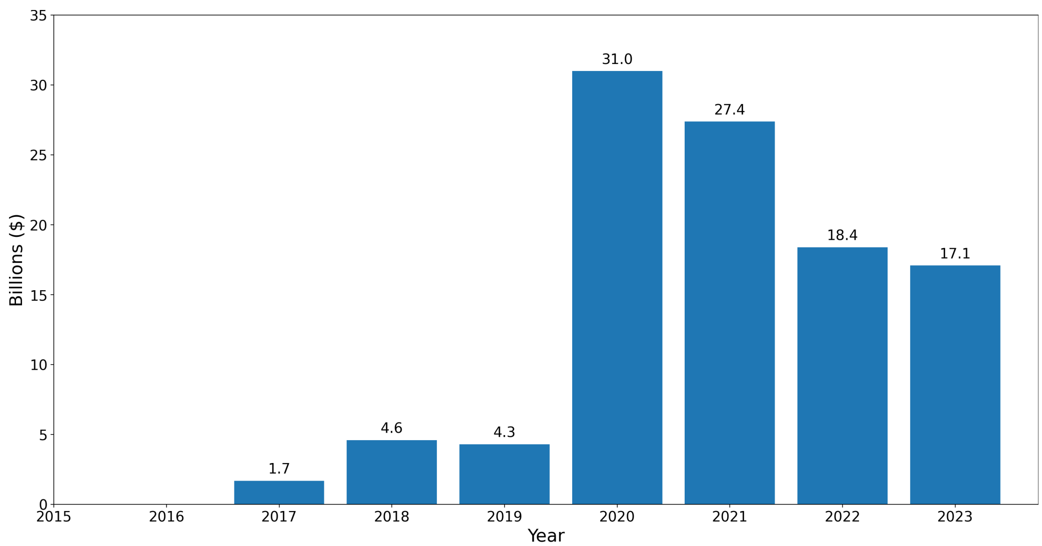
In the context of rebalancing towards these optimal portfolio weights, an individual would increase the allocation towards an asset that experienced a recent bad return (in order to bring the asset's portfolio allocation back towards its optimal weight). Conversely, an individual would decrease the allocation towards an asset that experienced a recent good return.

G Robinhood Growth post-2020

Figure IA.5 displays the growing number in funded accounts and net deposits on the Robinhood platform. The figure demonstrates that Robinhood continued to grow according to both metrics after 2020, indicating that the new class of retail traders does not appear to be a temporary phenomenon.



(a) Funded Accounts



(b) Net Deposits

Figure IA.5: Growth in Robinhood Funded Accounts and Net Deposits Over Time

This figure displays the number of funded accounts and net deposits on the Robinhood platform over time. Panel (a) displays the number of funded accounts on Robinhood over the years 2015-2023. A funded account is defined as a Robinhood account into which the account user makes an initial deposit or money transfer of any amount. Panel (b) displays the amount of net deposits into Robinhood accounts over the years 2017-2023. Net deposits include all cash deposits received from customers net of reversals, customer cash withdrawals and other equity and cash amounts transferred out of the platform. The data is from Robinhood's 10-K statements.

H Retail Brokerages Comparison

H.1 Individual Investor Literature Datasets

Table IA.19 provides a comparison of a subset of retail brokerage datasets previously used in the literature. The table includes datasets for the U.S., Sweden, and Finland. In contrast to the Robinhood dataset used in this paper, these previously studied datasets all cover periods during the 1990s. Moreover, while the median age of Robinhood investors is 31 years old, the average age of retail investors in these datasets is significantly higher (55 years for the U.S. dataset and 51 years for the Sweden dataset). The mean size of these accounts ($35,629 for the U.S. dataset and $34,595 for the Sweden dataset) is also significantly larger than that of Robinhood investors ($3,500). This demonstrates that the new class of retail investors represented by Robinhood traders is significantly younger and smaller than their previously studied counterparts.

Table IA.19
Individual Investor Trading - Prior Data Sources

This table provides a summary of previously used datasets to study individual investor behavior. Data includes a sample of research papers employing the dataset, region covered, time period covered, frequency of observations, number of investors, mean investor age, and mean investor size. Data is sourced from the relevant paper data sections.

Research Papers	Region	Time Period	Frequency	Investors	Mean Age	Mean Size
Odean (1998), Odean (1999), Barber & Odean (2000, 2001), Kumar (2009)	U.S.	1991-1996	Monthly	62,387	55 years	$35,629
Calvet, Campbell, & Sodini (2009)	Sweden	1999-2002	Monthly	4.8 million	51 years	$34,595
Grinblatt & Keloharju (2000, 2001, 2004, 2009, 2011)	Finland	Dec. 27, 1994 - Jan. 10, 1997	Daily	N/A	N/A	N/A

H.2 Major Retail Brokerages Worldwide

Key statistics for all current major retail brokerages worldwide are summarized in Table IA.20. The table lists the brokerage names along with the year in which they were founded, the regions they serve, the instruments that can be traded on them, the number of client accounts, the total assets under custody of non-discretionary client accounts, the average client age, experience level, and account size, and whether the brokerage is publicly traded. The table reveals that among the newer retail brokerages—those founded after 2000—Robinhood is the dominating force, with over $100 billion in assets under custody. The table also reveals that while users on the traditional brokerages such as Charles Schwab, Fidelity Investments, and Interactive Brokers tend to be somewhat older, more experienced, and larger in average account size, the same is not true for the newer brokerages. In fact, among the newer brokerages that service the U.S. and allow stock trading the average investor appears to have relatively little experience, small amount of capital, and be on the younger. This indicates that Robinhood investors are likely representative of the new class of retail investors.

Table IA.20
Summary of Major Brokerages Worldwide

This table lists the major retail brokerages worldwide along with their year founded, region(s) served, instruments traded, number of accounts, client assets under administration ($), average user age, average user experience level, average user account size, and whether the brokerage company is a publicly traded firm. Data is gathered from publicly available sources including brokerage 10-K statements where available, brokerage websites, and financial sites such as Investopedia, Motley Fool, NerdWallet, Trading Bible, and Compare Forex Brokers.

Brokerage	Year Founded	Regions Served	Tradable Instruments	Accounts	Assets (USD)	User Age	Experience Level	Account Size	Publicly Traded?
Charles Schwab	1971	Global	Stocks, ETFs, options, mutual funds, futures	34.8 million	$8.52 trillion	Mid-40s	Medium-High	$240,000	Yes
Fidelity Investments	1946	Global, primarily US	Stocks, ETFs, bonds, options, mutual funds, crypto	38.7 million	$7.7 trillion	18–35 (45% new)	Varied	Varied	No
Interactive Brokers	1978	Global	Stocks, ETFs, options, forex, bonds, crypto	2.4 million	$128.4 billion	35–50	Medium-High	$200,000+	Yes
Robinhood	2013	US	Stocks, ETFs, options, mutual funds, crypto	23 million	$102.6 billion	31	Low	$3,500	Yes
eToro	2007	Global	Stocks, crypto, ETFs, CFDs, indices	30+ million	$10 billion+	34–36	Beginner to Intermediate	$1,500–$3,000	No
Webull	2017	US, Hong Kong	Stocks, ETFs, options, crypto	13+ million	$8.2 billion	31–34	Intermediate	$2,500–$5,000	No
DEGIRO	2008	Europe	Stocks, ETFs, bonds, options	2.4+ million	$6.3 billion	35+	Experienced	N/A	No
XTB	2002	Europe, Asia, LATAM	Forex, indices, commodities, CFDs, ETFs	539,000	$410 million	30s	Intermediate	$1,000–$2,000	Yes
OANDA	1996	US, Europe, Asia-Pacific	Forex, CFDs	500,000+	N/A	30–45	Medium	Varied	No
Eightcap	2009	Global	Forex, CFDs, cryptocurrencies	N/A	N/A	25–40	Medium	$5,000–$10,000	No
Moomoo	2018	US, Hong Kong	Stocks, ETFs, options	5+ million	N/A	30–40	Intermediate	$5,000	No
SoFi Invest	2011	US	Stocks, ETFs, crypto, options	3+ million	N/A	25–35	Beginner to Intermediate	$2,000–$5,000	No
Public	2019	US	Stocks, ETFs, crypto	1+ million	N/A	25–35	Beginner to Intermediate	$1,000–$2,000	No
Stockpile	2010	US	Stocks, ETFs	100,000+	N/A	18–30	Beginner	$500–$2,000	No
Cobra Trading	2004	US	Stocks, options, short selling	N/A	N/A	35–50	Advanced	$50,000+	Yes