The Wisdom of the Robinhood Crowd

IVO WELCH*

ABSTRACT

Robinhood investors increased their holdings in the March 2020 COVID bear market, indicating an absence of collective panic and margin calls. This steadfastness was rewarded in the subsequent bull market. Despite unusual interest in some "experience" stocks (e.g., cannabis stocks), they tilted primarily towards stocks with high past share volume and dollar-trading volume (themselves mostly big stocks). From mid-2018 to mid-2020, an aggregated crowd consensus portfolio (a proxy for the household-equal-weighted portfolio) had both good timing and good alpha.

SSRN: https://ssrn.com/abstract=3696066
JEL: D9, G4, G11

Forthcoming: The Journal of Finance.

* Ivo Welch is from the Anderson School at Management at UCLA. This paper was written during the Covid lockdown. I thank many colleagues for informal feedback, especially Brad Barber and Terrance Odean not only for feedback but also for their sharing their data with me. I also thank Robinhood for making their data publicly available and Robintrack, created by Casey Primozic, for having collected it and making it available to researchers. Gabriel Cuevas independently replicated the paper and confirmed that the included code and data are sufficient to do so. Finally, I thank the referees, the Associate Editor, and the Editor (Stefan Nagel) for their constructive comments. I have no conflicts of interest to declare.

The online retail brokerage company Robinhood (RH) was founded in 2013 with a plan to make it easier and cheaper for small investors to participate in the stock and option markets. Its customers are widely believed to be a new breed of mostly young, small, computer-savvy but novice investor. Some reports have pegged their typical account sizes as low as $2,000. By mid-2020, Robinhood was reported to have attracted a clientele of over 13 million investors, more than twice the population of Finland.

The active participation of such tiny investors on this large a scale is a new phenomenon. Thus, the academic literature still knows very little about them. Moreover, even our knowledge about the holdings of larger retail investors is slowly becoming dated. The seminal work of Barber and Odean (2000) and Grinblatt and Keloharju (2001) uses data from the 1990s—a different era in a time before the Internet, social media, and low transaction costs.

It is difficult to imagine that small retail investors could have previously coordinated into a collective short-squeeze, as they did with Gamestop in January 2021. Gamestop, a brick-and-mortar retailer, had traded under $5 per share throughout the first half of 2020. It had then risen to about $20 per share on January 11, 2021 (representing about $5 billion in market cap). Somewhere along the line, the Melvin and Citron hedge funds had shorted it. As the subreddit chatroom *r/wallstreetbets* lit up with talk of punishing greedy billionaires, Gamestop's share price rose from $20 per share to $325 per share on January 29, 2021. At this point, RH halted further purchases of Gamestop. The frenzy ebbed off and the stock fell back to $225 per share on the next trading day, February 1, $90 per share on February 2, and $53 per share on February 4. In the meantime, many hedge funds had to liquidate large positions at huge losses. Melvin alone lost about $4 billion. Although it is impossible to attribute responsibility (because trading is anonymous), it is widely believed that RH investors played a central role. Andrew Left of Citron called them the "angry mob." Whether RH investors ultimately were or were not responsible, all short-sellers are now actively contemplating a world in which they could become exposed to such crowd risk at any moment.

In this paper, I raise the more basic question of what types of stocks RH investors tend to hold—that is, in more ordinary times. Taken together, are they spread over many holdings or concentrated in just some stocks? Related, would RH investors be better off not entering the market altogether? On the one hand, diversification into stocks may allow RH investors to improve their welfare. On the other hand, they may be cannon fodder for more sophisticated investors elsewhere. Only empirical analysis can tell us how their ventures into the stock market ended up.

From mid-2018 to mid-2020, RH offered an API that made it possible to query the number of (anonymous) RH investors (N_i) who held a particular stock (i) at that moment in time. In turn, the website Robintrack.net coded scripts to continuously pull down these holdings (at a speed of about 20 stocks per second, cycling through all stocks about once an hour) and reposted the data online with RH's blessing. Some of the more interesting patterns in these data are as follows:

(i) Changes in Returns and Holdings: Similar to larger retail investors in the 1990s (Grinblatt and Keloharju (2001), Barber and Odean (2008)), smaller RH investors in 2020 liked to purchase both large gainers and large losers.[1] Unfortunately, earlier samples did not include precipitous market declines, such as the market crash of 1987, the dot-com burst of 2000, or the financial crisis of 2008. In contrast, the 2018 to 2020 sample includes the COVID crisis. This makes it possible to investigate retail behavior during at least one sharp market-wide downturn and its quick subsequent recovery.

I find that aggregate RH holdings increased in this episode. Retail investors seem to like volatility, at least collectively.[2] During the sharp downturn, RH investors did not collectively panic or suffer margin calls. A first "purchasing" spike (i.e., an increase in the sum total number of investor holdings over all stocks) occurred as early as the day after large market changes (including drops). Presumably, this reflected the extant purchasing power in their accounts. A second spike occurred about four days later. This is roughly the time required to complete a bank transfer of funds to a RH account.

The evidence suggests that RH investors may have actively added cash to purchase more stocks. Thus, they collectively acted as a (small) market-stabilizing force. (Of course, because RH investors also bought more stocks after market increases, one should not consider them a stabilizing force in other situations.) Moreover, their behavior paid off. Coming into the market only gradually during an overall bull market, they earned a lower average rate of return than the stock market overall—but with timing good enough to earn a higher Sharpe ratio.

(ii) Level Holdings: Prior retail investor literature focused on the timing of trades. For some papers, this was the only perspective possible, because holding data were not available. In contrast, I am able to examine portfolio-level holdings.

[1] Barber and Odean (2008) suggested this could be due to stocks catching the attention of investors, while Grinblatt and Keloharju (2001) suggested it could due to deliberate sensation-seeking. See also Ben-Rephael, Da, and Israelsen (2017), DellaVigna and Pollet (2009), Fang and Peress (2009), Hirshleifer, Lim, and Teoh (2009), Peng and Xiong (2006), Da, Engelberg, and Gao (2011), and DellaVigna and Pollet (2009). Barber and Odean (2013) survey the behavioral literature on individual investors. Barber et al. (2020) investigate stock-specific RH holding changes in more detail.

[2] The data do not allow testing whether old RH investors panicked and were replaced by new RH investors with more risk appetite. In Seru, Shumway, and Stoffman (2010), investors stop trading upon poor performance but do not exit the market altogether.

I find that there is plenty of opportunity to ridicule RH investors. For example, they overweighted stocks that likely appealed to some unusual interests: Ford but not GM, Facebook in 2018 but not in 2020, and airline stocks in mid-2020 but not in 2018. AMD, Snapchat, and cannabis stocks were remarkably popular. At the end of January 2019, Aurora Cannabis (ACB) was briefly the most widely held stock, with 244,532 investors! (AAPL was second with "only" 237,521 investors.) Thus, RH-type investors may well have been playing a role in the active trading of, and the steady-state demand for, cannabis and some other (otherwise) obscure stocks.

Nevertheless, this "crazy mob" narrative is misleading. The typical RH investments were much saner. Cannabis and other "experience holdings"[3] were just minor sideshows—most investment weight was not in these stocks. This becomes obvious when one looks at a more representative *ARH* portfolio that holds $w_i \equiv N_i / \sum_i N_i$ in each stock, where i is a given stock and N_i is the number of RH holders in this stock. It is of course unlikely that any particular investor held this portfolio. Instead, *ARH* should be viewed as one reasonable "consensus statistic" among other viable statistics, akin to the notion of a consensus forecast, as in Zarnowitz and Lambros (1987). One can therefore view *ARH* as a "crowd wisdom" portfolio. In other data sets, an *ARH* portfolio also turns out to be a very good proxy for a household[4]-equal-weighted portfolio.

The simplest way to characterize how the *ARH* portfolio differs from a more conventional broad stock market portfolio is to describe its correlates. A "quasi-robinhood" (QRH) portfolio with two-thirds normalized share-trading volume and one-third normalized dollar-trading volume (both calculated over the previous year) has a 75% correlation with the *ARH* portfolio in investment weights. Its rate of return time-series correlation, after removing other prominent factor variation, is over 90%. Moreover, equivalent portfolios have the same correlations when using the Barber and Odean (2000) data. With correlations this stable and this high, this "quasi-retail" portfolio can even stand in as a proxy for retail holdings, at least for some purposes.

What is perhaps most surprising is that the *ARH* crowd portfolio did not underperform. This is the case for a zero-factor model (i.e., returns above the risk-free rate), a net-of-the-market model, a one-factor model (i.e., abnormal returns adjusted for market movements with beta), and a five-factor-plus-momentum model (Fama and French (2015), Carhart (1997)), which I refer to now as a six-factor model. The alphas of the *ARH* portfolio were

[3] I define an "experience holding" loosely as one in which an investor derives some pleasure from holding the stock in their portfolio, separate from a purely cash-flow-based perspective. My classification of certain stocks as experience stocks is plausible but subjective. A reader with a different assessment may mentally group stocks differently.

[4] In the Barber and Odean (2000) data, accounts are identified as households. I therefore adopt the same terminology, although the unit of observation could well be an account or investor instead of a household.

positive, and despite the very short sample, even statistically significant in the six-factor model, with a respectable abnormal rate of return of 1.3% per month. Thus, RH investors collectively were not cannon fodder exploited by more sophisticated investors elsewhere. Good timing and good stock performance can help explain why RH investors have continued to pour in.

Although the *ARH* consensus portfolio had good alpha (and low transaction costs), the data do not allow for an analysis of how well individual RH investors performed. I can speculate that individual RH investors probably held less diversified and thus riskier portfolios, but they may have actually *enjoyed* this risk.

The analysis in my paper focuses on investment holding *levels* rather than *changes*. From a performance perspective, this is reasonable. Holdings present a wider-angle picture than trades. Moreover, most investors' portfolio returns are more likely to reflect the performance of their long-held positions than that of their most recent purchases. Even at a monthly frequency, the median *absolute change* in *ARH* investment weight (over all stocks and months in the RH data) was less than 7% of the investment *level* weight. Thus, even among the clientele of RH investors, which is widely believed to be quite fickle, the *ARH* portfolio displayed strong hysteresis.

Yet the difference in my paper's focus also means that the portfolio performance results reported in my paper are not readily comparable to those in most earlier studies of retail investors. A new position may underperform right after the initial purchase but perform well over longer horizons. The *ARH* performance reported in my paper is more akin to (but not exactly the same as) the better-performing buy-and-hold retail investor portfolio used as the control benchmark in Barber and Odean (2000) than to the performance of the more active traders in earlier literature.

Moreover, even for the performance of retail *trades*, the literature is not in full agreement. Kaniel et al. (2012) and Kelley and Tetlock (2013) use proprietary trading data from the NYSE (2000 to 2003 and 2003 to 2007, respectively) and find that retail trades outperformed. Boehmer et al. (2020) find likewise using a novel metric for classifying (some) trades from 2010 to 2015 as retail trades. But Barber and Odean (2000, 2002), Grinblatt and Keloharju (2000), Barber et al. (2020), and others find that active trades by retail investors underperformed. Possible explanations for this discrepancy is that performance could depend on the holding interval (Barber and Odean (2008)) or broker (Fong, Gallagher, and Lee (2014))—or more simply that historical return performances of investment strategies have low external validity in general.

Other academic studies of RH investors are also now emerging. Moss, Naughton, and Wang (2020) show that RH investors did not care much about socially responsible (ESG) investing, contrary to some experimental studies. Barber et al. (2020) find that *herding-related* buying was not advantageous, losing as much as 5% over five days.[5] Moreover, they show that during RH outages, trading in stocks dropped by 0.7% (6% of retail activity), and even more for the 50 most popular RH stocks. Ozik, Sadka, and Chen (2020) show that RH investors traded more when COVID lockdowns took effect, effectively "grounding" more investors in front of their computers at home. Ben-David et al. (2020) identify RH investors as "sentiment driven" to classify ETFs by appeal to cost-conscious versus sentiment-driven investors.

Furthermore, the *ARH* investment weight is a monotonic transform of the "breadth of ownership" measure. While Chen, Hong, and Stein (2002) find that stocks held by fewer mutual funds subsequently underperformed, Nagel (2005) finds that this relationship disappeared in a sample that included five more years (according to Choi, Jin, and Yan (2013), who investigate *changes* in breadth of ownership). Again, even statistically significant return performance, carefully researched, often does not hold out of sample.

The paper is organized as follows. Section I describes the RH brokerage firm, the Robintrack data, and caveats about what the data can and cannot tell us. Section II shows how individual stock volatility has increased RH investor interest. Section III examines the effects of aggregate market movements. Section IV switches to and characterizes the *ARH* portfolio and its "Q" factor proxy. Section V describes portfolio performance, and Section VI assesses transaction costs. Finally, Section VII concludes.

I Background

A. Robinhood

Wikipedia describes Robinhood Markets, Inc. as a Financial Industry Regulatory Authority (FINRA), U.S. Securities and Exchange Commission (SEC), and Securities Investor Protection Corporation (SIPC) regulated retail broker-dealer headquartered in Silicon Valley. It has no brick-and-mortar presence, and handles its (mostly small) retail customers through a mobile app and a website. It allows investing in stocks, ETFs, options, and crypto currencies. RH also has had its fair share of controversies, related to service outages, banking licenses, an SEC probe, insufficient disclosure of payment for order flow, a security breach, the trading halt during the 2021 Gamestop event, and even the suicide of one of its investors.

[5] A nonpublic J.P. Morgan research report by Cheng, Murphy, and Kolanovic (2020) studies RH changes. Unlike Barber et al. (2020), they emphasize good timing of RH investors *on average*. Their finding that RH investors purchase strong winners and losers overlaps with the findings in Section II below.

RH's operations are largely like those of other retail brokerage firms. This also means that much of RH's internal business model remains obscure. It earns its own revenues through margin fees, cash balance interest, and payment for order flow (through market-makers such as Citadel or Virtu). In early 2020, RH was reported to have had about 13 million users. By August 2020, RH had raised another $200 million of fresh capital, boosting its valuation to $11.2 billion. By February 2021, RH had raised a further $3.4 billion. It IPO'ed on July 28, 2021, at $38/share, pegging its valuation at $32 billion. Soon after the IPO, it reached a peak of $70/share on August 4, 2021 before falling back to about $45/share.

Information about RH's customers has been leaked only sporadically. The Wall Street Journal reported on September 12, 2020 that "According to Robinhood...first time investors accounted for 1.5 million of its 3 million funded accounts opened in the first four months of 2020." The website Brokerage-Review.com estimated that the average account size at RH was only $2,000. Although RH investors are individually quite small, they are nevertheless collectively often seen as the "future of investing."

RH has been offering its customers many small technological innovations, such as a friendly mobile-first user interface. However, one important appeal is the commission-free trading.[6] From the perspective of its customers, RH may be spreading the now-eliminated (fixed-cost) fee into higher (per-share) variable costs, making it cheaper to buy and sell small positions (and even fractional shares), but potentially more expensive to buy and sell larger positions *if* other retail brokers were to transact better. (It is not known whether other brokers execute better or not.)

[Insert Table I here]

Table I collates public information about payment for order flow and implied trading volume for equities.[7] Impressively, RH more than tripled its trading volume in one quarter, growing from about 2% of the U.S. market in Q1 2020 to about 5% in Q2 2020.

B. Robintrack

Robintrack.net (RT) was created in 2018. For about two years, from May 5, 2018 to August 13, 2020, RT ran scripts to download the data made publicly available on RH's API. RH then terminated this public API and RT froze its operations. By this time, the

[6] RH's business model and rapid growth are widely viewed as having been disruptive in the industry. For example, other brokers also abandoned brokerage fees by October 2019, and Charles Schwab and TD-Ameritrade merged in mid-2020 to meet the competitive threat.

[7] RH earned just under twice this amount from the more lucrative options trading.

database had accumulated 3.5GB of data. After removing repeated intra-hour observations and unchanging holdings, it contains about 12 million ticker–hour observations. For each stock, I extracted the prevailing number of investors for each stock as of the last UTC-stamped data point on each day.[8] This procedure resulted in 5,777,002 RH ticker–day observations for 802 unique days and 8,597 useful tickers.[9] Of the 8,560 RH tickers, 8,387 were matchable to CRSP (the University of Chicago stock price data base). Of these, 3,834 were ordinary equity (CRSP sharecode 10 or 11). My paper limits attention to this specific ex-ante identified investment universe, that is, it does not include options and cryptocurrency holdings.

Some tickers do not appear at all, others appear late in the RT data. Early versions of the script probably omitted dual-class tickers ending with *.A* or *.B*, most prominently Berkshire Hathaway. RT remedied this with an upgrade on January 16, 2020.[10] Nevertheless, some stocks are never found in the RT data set for unknown reasons, most prominently CELG (Celgene) and TWX (Time-Warner).

Over time, the number of RH investors increased, and given the law of large numbers, presumably so did the reliability of the number of RH investors holding individual stocks. My paper focuses principally on the sample from June 1, 2018 to August 13, 2020 which contains 546 valid CRSP trading days. RH itself suffered some systemwide outages on March 2, 2020 and March 9, 2020. The RT script failed to run on August 9, 2018, January 24 to 29, 2019 (four days), and January 7 to 15, 2020 (seven days). The last RT outage coincided with the dual-class script update. Unfortunately, my data end before the infamous Gamestop episode.

C. Caveats

It is important to offer appropriate cautions before the analysis. The first major caution relates to external validity. The behavior of RH investors may or may not be representative of the behavior of other retail investors in 2020—much less the behavior of retail investors in another U.S. brokerage firm 30 years ago, Finnish investors 25 years ago, retail investors one decade ago, or retail investors one decade into the future. The period studied here is unique. It includes a market-wide transition to a zero-fee brokerage, the COVID 2020 bear market and recession, and the subsequent bull market. Extreme events (for example,

[8]The NYSE closes at 4:30pm EDT, which is either 22:20 or 23:30 UTC.

[9]The RT data were *not* professionally maintained and cleaned and thus require extra care. In particular, it contained some incorrect tickers, such as _OUT, _PRN, MTL-, PKD∼ (and its sibling PKD), which I hand-removed.

[10]Thus, BRK.B suddenly appeared with 38,023 users (and BRK.A with 134 users). (Figure A1 shows that BRK.A was right on the fitted line on June 30, 2020.) Other noteworthy dual-class examples that also appeared on January 16, 2020 are Royal-Shell Dutch and Lions Gate Entertainment.

2020, just like 1929, 1987, 2000, or 2008) are by nature near-singular and have unique aspects. However, by virtue of being extreme events, they can also provide us with a better glimpse into investor behavior—during normal and stressed times—than what is otherwise possible.

As in all empirical studies, each reader must make her own subjective assessment as to which results are likely to have external validity. My own assessment is that investment behavior does have good plausible external validity (and some of this was indeed subsequently verified in the Barber and Odean (2000) data), while stock return performance does not. In my view, the return performance of *any* investment strategy is not easy to extrapolate, regardless of publication venue. Nevertheless, the historically realized return performance in this context is interesting, as it can help explain why RH investors have been continuing to pile in.

The second major caution relates to the limited information available in RT. Without access to more disaggregated investment holdings and transactions of individual investors, it is only possible to investigate aggregate and relative stock-specific holdings (and holding changes). This obscuration of the data interacts with two more concerns. First, RH does not support ACAT transfer of existing brokerage accounts. Thus, the sudden simultaneous increases of holdings long after purchase due to funding transfers are not a concern.[11] Second, RH gives each investor a prize: one free randomly chosen share upon signup or referral. Among these prizes, 2% are "winners" because they receive one share drawn from six stocks priced above $10 and explicitly named on RH's bonus page. The shares given to the other 98% are impossible to ascertain.[12] Although investors could sell their initial share, and even with zero commissions and fees, it is likely that many investors simply hold on to it. With such holdings representing only one single share, they could seem more important than they are. Unfortunately, this is impossible to investigate or control for.

In sum, the number of investor holdings is a reasonable but noisy proxy for actual RH investments. An absence of findings could have been interpreted as due to insufficient power in the presence of noise, but low statistical power does not seem to be a problem. Moreover, below I show that the findings also hold in the Barber–Odean data, where complete information can be traced without obscuration noise or share awards.

[11] To fund their accounts, millions of RH investors handed their bank user account credentials (*including passwords*) to an intermediary named "Plaid," which then linked RH to their bank accounts. If Plaid were to be hacked, banks would assume no liability for withdrawn funds.

[12] A survey of past recipients confirms that the selections did not adhere to a literal reading of the description in the bonus offer. That is, these shares were not drawn from the three highest-capitalized stocks with prices below $10. While the exact mechanisms should not matter to RH investors, it could matter for some academic studies. I suspect that RH merely reassigns some random share just sold by another investor to the new investor, thereby saving external costs. If this is correct, the sale of one larger position could splinter into more holders rather than fewer—though still in line with new investors signing up.

II Stock-Specific Responses To Large Returns

I begin my investigation by examining whether firm-specific stock returns can predict subsequent *changes* in RH holdings. This information will be useful when assessing whether the evidence in the next section on aggregate purchasing patterns generally fits with investors' trading preferences.

Grinblatt and Keloharju (2001) show that Finnish investors from 1994 to 1997 purchased on either up or down movements, and Barber and Odean (2008) show that U.S. investors did likewise from 1991 to 1999. It is not a foregone conclusion, however, that RH investors act in similar fashion, because neither the types of investors nor the time period are the same. In particular, RH investors in my study are smaller and trade 25 years after the 2000 tech collapse in a period with more volatility.

A. Some Extreme Holding Changes

Before proceeding to the statistical analysis, Tables II and III provide an intuitive characterization. They describe the most extreme cases of changes in *ARH* investment weight, that is, in a portfolio formed in accordance with the number of RH investors holding each stock (as described in the introduction). These weights are investigated in more detail in Section IV .

[Insert Table II here]

Table II lists stocks with unusual one-day increases in RH investor interest. The common feature seems to be highly unusual one-day stock price increases or decreases. Some events also involve stock splits. Note that stock splits do not mechanically change the number of investors in the stock.

[Insert Table III here]

Table III lists stocks with unusual one-day decreases in RH investor interest. One again finds some rather stark stock price changes, in this case tilting more towards losses. Sounding somewhat trite, investors in India Globalization (IGC) appear to have decided to realize their profits after stark price increases. (IGC leases heavy equipment in India, develops cannabis-based therapies in the U.S., and runs a hotel in Malaysia.) However, the stock price patterns in these cases of extreme declining interest were not as strong.

B. Purchasing Individual Stocks on Large Price Changes

[Insert Figure 1 here]

In Figure 1, all trading days are first categorized by their rates of return net of the market (both from CRSP). The plots then tabulate the RH change statistics on the following trading day. The figure shows that RH investors purchased individual stocks when their prices increased or decreased greatly. This result is consistent with the attraction of attention in the Barber and Odean (2008) sense and/or sensation-seeking in the Grinblatt and Keloharju (2001) sense.

The analysis makes no attempt to distinguish between causal and correlative-only associations. It could be that other news first induced an instant stock price change that was followed on the next day by investor trades (possibly ignorant of the recent price change itself). The time precedence, however, ensures that later RH trades themselves would not have induced the stock return to move in the first place (as in Barber, Odean, and Zhu (2008)).

Panel A of Figure 1 plots the fraction of stocks that experienced increases in their RH holdings versus decreases in their RH holdings. It shows that when stock prices increased or decreased by about 20% (in logs) on one day, the number of RH investors increased *on the following day* in roughly three out of five cases (30% more buys than sells).[13] These strong increases in RH holdings contrast with small RH decreases after stock–days with small nondescript stock returns. The panel also plots (in lighter-colored lines) results based only on pre-2020 data and post-2020 data. The observed pattern was stable over time, if anything becoming stronger.

Panel B of Figure 1 plots *ARH* portfolio weight changes. Changes in these weights account for the fact that (a) an increase from 50 to 51 holders is not the same as an increase from 1 to 100 holders, and (b) other stocks may also have experienced changes on the same days. This following-day response still has a U-shape, but the panel shows that the increase in RH investors was stronger on the upside than on the downside. Taken together, the two plots suggest that an increase in interest after a downturn often occurred among smaller stocks, where just a few net purchasers would have increased the total number of RH investors.

[13]This is only about one-third of the effect, because the number also increased in approximately four out of five cases on the same day of the large price change itself. Intraday investigations show that RH investors respond to rather than anticipate large price movements. The same larger intraday effect also appears in Panel B. Because firm-specific purchases and sales are not the focus of my paper, this intraday evidence is omitted. Of course, it also does not seem plausible that RH investors would have much higher intelligence than other investors. The first-order source of the same-day correlation "should" therefore logically be subsequent trading rather than preceding knowledge.

In sum, the evidence above suggests that RH investors purchased stocks after large price movements, just like their larger counterparts 25 to 30 years earlier. This effect is weaker for large stock price decreases than for large stock price increases.

III Aggregate Holdings and the March 2020 COVID Decline

The more novel situation is the stark prolonged market-wide downturn. Retail investors could start to panic, or they could face margin calls. Both Grinblatt and Keloharju (2001) and Barber and Odean (2008) were limited in their ability to investigate such behavior, because there simply was no large decline in the Grinblatt–Keloharju sample, and the biggest decline in the Barber–Odean sample was only 15% (July 1998). In contrast, the 2018 to 2020 sample contains the COVID bust and boom. After reaching a high of 3,386 on February 19, 2020 and falling back to 3,130 on March 4, 2020, the S&P 500 fell to as low as 2,237 on March 23, 2020. This represents a 33% decline from its high a month earlier, and, perhaps even more unusual, with reasonable expectations of a calamitous economic depression caused by the COVID pandemic looming ahead. Remarkably and perhaps unexpectedly, although the economic depression did materialize, the stock market recovered to 3,000 by the end of May and proceeded to reach an all-time high of 3,662 on December 1, 2020.

[Insert Figure 2 here]

Visually interpreting the aggregate time-series behavior in response to stock returns is mildly complicated by the fact that RH experienced strong investor growth throughout the entire three-year sample period. Figure 2 shows a full-sample tally of the number of holdings.[14] The sum total of RH holdings grew steadily by 0.22% per trading day (77% per year) from mid-2018 to the end of 2019, then dramatically accelerated to a peak growth rate of about 3% per trading day around the end of March 2020, and finally slowly decelerated back to (around) its original 0.22% growth rate. Thus, the steepest RH growth was roughly contemporaneous with the COVID onset and stark decline in the stock market. Indeed, Ozik, Sadka, and Chen (2020) plausibly suggest that the COVID lockdown itself may have contributed to this acceleration.

[14] A smaller spike in January hints at tax-related, bonus-related, or New-Year-resolution-related purchases. The plot in Appendix Figure A1 provides a closeup view.

[Insert Figure 3 here]

Figure 3 presents the same information but for different horizons in an x-y graph. The figure plots the performance of the S&P 500 and the contemporaneous percentage change in RH sum total holdings. The measuring intervals for both variables are equally long and consecutive intervals are overlapping. The figure shows that over horizons from one day to one month, both large decreases and large increases in the S&P 500 are associated with large contemporaneous increases in the sum total number of holdings by RH investors.

Although one should not attribute the inflow of RH investors to deliberate investment planning, one can ask how well RH investors happened to have timed market entry. The lowest number of holdings was 4.3 million on the first day of my main sample (June 1, 2018), while the highest was 43.1 million on the last day of my sample (August 13, 2020). Imagine that every holding represented an equal investment, that funds not invested in RH holdings were in cash, and that funds allocated to RH were in the S&P 500. For convenience, assume that the rate of return on cash is zero. Under these assumptions, on June 1, 2018, RH investors would have held 13% of their funds in the market and 87% in cash. Over the following 545 trading days, their equity allocation would have gradually increased, with an average market holding of about 35%, reaching 100% only on the final day, August 13, 2020.

The investment-weighted rate of return for such investors, based on (one-day-lagged) holdings, would have been 3.57 basis points (bp) per day (19.6% over the entire sample). Investors who always held the market would have earned 5.14 bp per day (28% over the entire sample). However, the standard deviation of RH investors would have been only 0.76% per day, while the standard deviation of the market would have been 1.6% per day. Thus, the Sharpe ratio of RH investors would have been higher than that of the stock market overall. Though unlikely to be deliberate, the macro timing of RH investors turned out to have been lucky.

[Insert Table IV here]

Table IV statistically investigates the association between daily percentage changes in the S&P 500 and daily percentage changes in the aggregate number of RH holdings. The two left regressions use the entire sample. The two right regressions begin in February 2020 (thereby omitting the January 2020 RT script update). In explaining the inflow of RH investors, the S&P 500 return has to compete with both a constant and a battery of lagged autocoefficients, which attribute various trends to unrelated factors. These controls also capture the acceleration of RH holdings at the onset of the COVID crisis. Some specific lags have no particular reason to be in the regression, except as placebos to confirm that this is not the kind of regression where everything comes out significant.

The estimated coefficients on S&P 500 market returns suggest a first spike in RH holding changes on the day of and one day after the stock market increased or decreased. This mirrors their behavior after individual (not aggregate) stock price increases or decreases, as described in the previous section. This timing suggests that some RH investors learned about strong market movements towards the end of the day and added further holdings via their existing (margin) purchasing balances on the following day.

More intriguingly, a second investment spike occurred about four to five days later—roughly the time required to complete a bank transfer. This second spike of about 9% on the bear side in the post-January 2020 regressions suggests that the 33% COVID market drop would have increased the aggregate holdings of RH investors by about 3%. This increase would have likely been funded by dollars previously held in checking or savings accounts and would have now flowed into the riskier equity markets.

Note that although RH investors' response to stock market changes is quite respectable and statistically significant, these stock market changes can explain only a tiny part of the nearly tenfold increase in RH holdings. Depending on the column, the coefficients on either positive or negative market movements add up to only about 0.1 to 0.2. Thus, even the 30% increase in the overall stock market during the sample period could not have been a major driver of RH holding increases and/or signups.

In sum, just as strong price increases or decreases can predict additional RH holdings of individual stocks in the cross-section, strongly positive or strongly negative overall market movements can predict increases in RH positions in the time series. This evidence suggests that RH investors did not retreat from the stock market during its starkest decline. Instead, it appears that they played a small but active market-stabilizing role during the COVID crash. (They are not necessarily a stabilizing force in general, because they also buy aggressively in bull markets.) In a study of the selling behavior of mutual funds and hedge funds and the purchasing behavior of RH investors, Glossner et al. (2020) reach a similar conclusion. Harms (2021) also finds similar behavior among young German and Austrian investors.

In this case (as in all other U.S. stock market crashes in our lifetimes), holding or increasing one's market exposure after a market bust would have paid off handsomely in the subsequent stock market boom.

IV The *ARH* Portfolio Perspective

In the aggregate, RH investors had fortunate timing for their entry into stock market investing. This raises the question of how their holdings performed in the cross-section. Did they buy the right stocks?

In the introduction, I defined the *ARH* crowd portfolio as

$$w_{i,t}^{\text{ARH}} \equiv \frac{N_{i,t}}{\sum_i N_{i,t}} \,, \tag{1}$$

where i and t index stocks and trading days. The *ARH* portfolio invests more when more investors are holding a stock. For example, if there are twice as many owners of stock A compared to stock B, the weight of A in the *ARH* portfolio is twice that of B. Note that the price of the stock is irrelevant. Thus, an investor with some number of shares in a \$1 stock and some number of shares in a \$100 stock is considered to have placed 50% weight on each stock. The weight $w_{i,t}^{ARH}$ would therefore be more representative of two investors with 100 shares in the former and one share in the latter (i.e., investing \$100 in each stock) than of two investors with 50 shares in each (i.e., investing \$50 in the former and \$5,000 in the latter).

A. ARH *As Household Equal-Weighted Proxy*

ARH could plausibly proxy for a household-equal-weighted (HH-EW) portfolio if investors hold roughly equal-sized (or zero) positions proportional to their wealth. Unfortunately, data limitations prevent us from tracing RH investors' actual holdings. Fortunately, Brad Barber's and Terry Odean's generous sharing policy made it possible to investigate equivalent portfolios in the Barber and Odean (2000) data. Their data contain actual month-end portfolio holdings, by account, at a smaller discount brokerage firm from 1991 to 1996.

Panel A in Appendix Table A.I shows that an *ARH*-equivalent portfolio (constructed using the number of investors in each stock and analogously named *ABO*) had a 97.1% correlation in investment weights with the true HH-EW portfolio. It also shows that the equivalent correlation for the trading-volume-based *QBO* proxy portfolio, investigated below, is lower, at only 72%.

Panel B switches to a time-series rate of return perspective. The regressions are run on 59 months of data. The panel shows that the correlation between the true full-information Barber and Odean (2000) portfolios and their information-obscured proxies. The first two lines describe raw returns, the second two lines first remove the Fama and French (2015) five factors from the time-series portfolio returns before calculating correlations.

The correlations between the *ABO* and HH-EW portfolio are nearly perfect. (The association with the trading-volume-based *QBO* portfolio is still a very solid 84.3%, suggesting that it can also sometimes serve as a proxy for retail trading.)

B. Alternative Crowd Portfolios

The *ARH* portfolio is not the only crowd portfolio that could be considered. One of its drawbacks is that when a stock increases in value, its tilt in the *ARH* portfolio increases only within the rebalancing interval, not across the rebalancing intervals. (To address this concern, the tables below also show return performance results for longer holding periods.)

In an alternative crowd portfolio, each investor (holding) could represent an equal number of shares. In this case, $w'_{i,t} \equiv (n_{i,t} \cdot P_{i,t})/(\sum_i n_{i,t} \cdot P_{i,t})$, where P_i is the prevailing stock price. The first drawback of this alternative crowd portfolio is that any variable correlated with price (such as market cap, dollar-trading volume, or simply price itself) would become nearly mechanically correlated with this portfolio's investment weights. That is, the portfolio investment weights would no longer be based only on information obtained from RH but also on information from CRSP. The second drawback is more pragmatic and shown in Appendix Table A.I. The "ABO×P" portfolio is not as good a proxy for the HH-EW portfolio as the *ABO* portfolio.

In sum, I consider an *ARH* portfolio that is one feasible crowd wisdom portfolio. Its investment weights summarize RH crowd participation in a way that is likely very similar but not identical to the investment weights in a HH-EW aggregate portfolio.

C. Odd but Unimportant Holdings

[Insert Table V here]

The *ARH* portfolio is peculiar enough that it is important to describe it more intuitively before analyzing it statistically. In fact, it is so peculiar that it occasionally provokes public ridicule. RH investors overweighted some rather unusual portfolio positions.

For example, the *ARH* weight of India-Cannabis (IGC) was much higher than its value-weighted market-cap-based weight (from CRSP). Table V shows that IGC ranked 27th in the *ARH* holdings in 2019 with 0.59% of the *ARH* portfolio. *Yahoo!Finance* describes this 50-employee company as follows:

> *India Globalization Capital, Inc. purchases and resells physical infrastructure commodities. The company operates through two segments, Infrastructure Business, and Life Sciences. It buys and sells infrastructure commodities, such as steel, wooden doors, marble, and tiles; rents heavy construction equipment, including motor grader, transit mixers and rollers; and undertakes highway construction contracts. The company also develops cannabinoid-based products and therapies, such as Hyalolex for the treatment of patients from anxiety, agitation, dementia, depression, and sleep disorder diseases; and Serosapse for the treatment of Parkinson's disease. In addition, it offers offer extraction, distillation, tolling, and white labeling services under the Holi Hemp brand; and hemp crude extracts, hemp isolates, and hemp distillates. The company operates in the United States, India, and Hong Kong. India Globalization Capital, Inc. was founded in 2005 and is based in Potomac, Maryland.*

It is difficult to imagine a rational portfolio in which IGC would deserve an investment weight similar to that of J.P. Morgan.

Although not shown in the table, other cannabis stocks also attracted unusual interest from RH investors. (The RH mobile interface even has a special button dedicated to displaying the most popular cannabis stocks.) As already noted in the introduction, at the end of January 2019, Aurora Cannabis (ACB) was briefly the most popular stock on RH.

Table V further shows that RH investors had a relative holding (and trading) preference for certain oil&gas exploration and biopharmaceutical-related companies. The common denominator seems to be that these stocks had high idiosyncratic risk.

In sum, it is easy to spin a tale in which RH investors fit the stereotype of unsophisticated gamblers who buy on impulse and earn low returns—being taken advantage of by more sophisticated professional traders. Indeed, Barber et al. (2020) show that *during a few extreme herding episodes*, some ex-ante identifiable trades were so ill-timed that they lost 5% immediately (although Cheng, Murphy, and Kolanovic (2020) report good timing *on average*).

D. Important Holdings

However, the cannon-fodder narrative above is incomplete to the point of being misleading. It is true that RH investors were greatly overrepresented in some rather unusual stocks (as in IGC) and suffered from all sorts of behavioral and perhaps sometimes harmful patterns when timing their buys and sells. But this does not mean that the most important holdings in their consensus portfolio were in these rather unusual stocks. Put differently, the unusual holdings described in Table V—though possibly of real (and distorting) consequence to a number of small companies—were not the rule but rather the exception. From the perspective of the overall *ARH* portfolio, cannabis stocks can be described as interesting but small experience or curiosity holdings.[15]

[Insert Figure 4 here]

Figure 4 provides a graphical perspective of *ARH* portfolio holdings from two snapshots, one at the end of 2018, and the other towards the end of the sample in mid-2020. The blue line is a smoothed version of the perfectly monotonically declining relation between the log market-cap rank and its weight in the market-cap-weighted portfolio. The black line is a smoothed version of the relation between the log market-cap rank and the weight in the *ARH* portfolio.

Although IGC was perhaps the strangest popular holding in Table V, it was not even particularly noteworthy in the figure. After all, even at its height, IGC still represented only 0.5% of the portfolio, leaving 99.5% for other holdings. Many other stalwart stocks represented more than twice IGC's weight. The other outliers from Table V are not even noticeable in the cloud of other holdings.

This does not mean that the *ARH* portfolio mimicked the market cap value-weighted portfolio either. For example, the top plot for December 31, 2018 shows that Ford (F) was nearly as popular as Apple (AAPL) and Microsoft (MSFT), with 3.5% versus 3.8% and 2.9%, respectively. Yet Ford's market cap-based weight was much lower, with 0.13% versus 3.3% and 3.4%, respectively. Other popular stocks were AMD, Fitbit, GoPro, Netflix, and Snapchat—all firms selling products familiar to computer-savvy younger customers.[16]

Comparing the top and the bottom plots shows that the (blue) market-cap-weighted curve steepened over the sample period, with AAPL reaching a value of $1.6 trillion and representing about 5% of the overall stock market. In contrast, the slope of the black RH curve for *ARH* market cap weighting remained roughly unchanged. Although the *ARH* portfolio also invested strongly in AAPL (with about 2% weight), it did so less aggressively than the value-weighted market by mid-2020. Instead, RH investors overweighted Disney, GE, Ford, and airline stocks. (Remarkably, they did not hold unusually large investments in Tesla, perhaps the most surprising large gainer during the COVID crisis. This suggests

[15] My analysis understates the "sanity" of *ARH*, because it excludes many better-diversified ETF portfolios (with CRSP share codes other than 10 or 11). These are also commonly held by RH investors but are not considered in my study. Like most other studies of retail customers, I also have no data on their external investments.

[16] It seems unlikely that RH attracts the same lower socioeconomic clienteles that were studied in Kumar (2009). However, RH investors do seem to share some of the gambling predilections of these clienteles.

that retail investors were not especially involved during Tesla's quintupling.) A fair characterization is that although retail investors loved "new-economy" stocks in mid-2018, their portfolio tilted more towards fallen "old-economy" stocks by mid-2020.

[Insert Table VI here]

Table VI lists the largest holdings by RH portfolio investment weight. With data for both snapshots for a given stock, it provides an insight into the evolution of these holdings. A stock can see increases both in its ranking relative to other stocks and in its market cap-relative weight in the RH portfolio if

(i) it is *more* actively purchased by (new) investors than other stocks, and

(ii) its price (and thus its weight in the value-weighted market portfolio) decreases compared to other stocks.

The table shows instances of both. For example, RH investors flocked to American Airlines (AAL) although or perhaps because AAL had dropped from a high of about $50 per share in late 2018 to about $10 per share in mid-2020 (reducing its weight in the value-weighted CRSP index from about 0.06% to 0.02%). The 7,300 original holders in 2018 grew to 654,611 holders in 2020, far beyond the average 360% increase in the total number of RH holdings. AAL's weight in the *ARH* portfolio thus increased from 0.12% to 2.39%. A quick peek at its stock price in February 2021 suggests that this turned out to be a good choice, with AAL now trading at $17.27 per share. Facebook (FB) also had good stock market performance, raising its weight in the value-weighted market from 1.34% to 1.78%. It fell out of relative favor with RH investors, however, because FB's holders at RH increased by only 40% rather than by the 360%. PLUG (a hydrogen fuel researcher) benefited from both strong rate-of-return performance and increased interest.

In sum, even though some of the investments in small and unusual experience stocks are eye-popping, these positions represented only a small part of the RH crowd portfolio—they were a storm in a teacup. The big picture was in larger stocks, typically more familiar to (and some focused on) younger retail consumers.

E. *Mimicking The* ARH *Crowd Portfolio*

It is easiest to explain the nature of the *ARH* portfolio by describing its correlative determinants.[17] A few issues should be kept in mind when doing so.

[17] (Market-wide relative) trading volume in individual stocks is not exogenous. There is no empirical evidence that an engineered deliberate exogenous shock to trading volume (whatever this might mean for an endogenous variable in the first place) would change RH holdings.

First, it is more important for a prediction of investment weights to explain the top 50 holdings than the bottom 3,000 holdings. These top holdings are by and large not unusual experience micro-stocks like IGC. Mispredicting the weights of cannabis stock is forgivable; mispredicting the weight of AAPL is not. Second, like most holding portfolios (and especially aggregated holding portfolios), the *ARH* portfolio exhibits strong hysteresis. Figure 4 and Table VI show that many investment weights were stable over more than a year. Thus, short-term variables (like recent rates of return) are unlikely to help explain much of the investment weights. Third, it is more useful to explain *ARH* if the result can be used to create a proxy for *ARH* beyond the sample period in which we already have *ARH* itself. Thus, it can be more interesting to understand retail holdings in relation to easily available variables than in relation to boutique variables.[18]

With 1.7 million matched ticker–day investment weight observations, it is feasible to start with minimal theory and try to disentangle similar variables empirically. A 50-variable "kitchen-sink" regression—containing variables such as IPO date, alphabetic index of name and ticker, market cap, minimum or maximum to highs and lows, share prices and returns, index membership, financial statement variables, and various holding periods and transformations—could achieve a correlation between predicted and actual weights in the *ARH* portfolio of about 80%.[19]

However, 50-variable kitchen-sink regressions give little intuition and usually have little stability. When explaining *ARH*, the most highly correlated portfolio was one constructed using only *share-trading volume* over the previous 12 months. This portfolio has a 70% weight correlation with *ARH*, only about 10% less than the full 50-variable kitchen-sink model. The next-highest weight correlation is for *dollar-trading volume*, even including share-trading volume. Combining the two trading volume variables yields an investment weight correlation of about 75%—only 5% lower than the full 50-variable kitchen-sink model.

Interestingly, these two volume-based variables are also plausible measures of retail *trading*, though not necessarily of retail *holdings*. (It is not a foregone conclusion that retail investors would be the ones ending up holding highly liquid volume stocks.) Interestingly,

[18]The attention measures in Lee and Ready (1991), Kelley and Tetlock (2013), and Boehmer et al. (2020) track retail investor order *flow*. These measures are likely to be related more to *changes* in investment weights than to investment portfolio *levels*. Their main drawback is that they are likely to remain boutique variables because they are not easy to obtain and replicate.

[19]I can speculate that the remaining $(1-0.8^2)\approx 35\%$ unexplained variation in investment weights relates to stock attributes for which there are no readily available (long-term) comprehensive proxies. Such variables might include retail customers' excitement about products, measures of the share of consumer expenditures spent on these companies, companies' brand values and advertising, the excitement and promise of new technologies, and/or perhaps even product introductions and recent performance. These variables could perhaps help explain why RH investors preferred Delta and American Airlines over United Airlines, Ford over General Motors, Disney more in 2020 than in 2018, or Facebook more in 2018 than in 2020.

neither the rate of return over the last year, nor the volatility, nor the price peaks or troughs add much economic explanatory power to this dual-volume model. Although stock returns do appear to prompt retail purchasing, these purchases are already sufficiently summarized by trading volume.

[Insert Table VII here]

More specifically, the two trading-volume metrics are best combined by first transforming each variable into a portfolio weight ($w_i(v_i) \equiv v_i / \sum_i v_i$). Panel A of Table VII shows that the individual metrics alone are not as good in explaining the *ARH* weights as their weighted average (*QRH*). This weighted *QRH* average, in rough accordance with estimated coefficients, assigns twice the weight on share volume as on dollar-trading volume.[20,21]

The statistical model was fit only to explain the *ARH* investment weights with *QRH* weights. It ignores the rate-of-return fit. Panel B shows the resulting contemporaneous rate-of-return correlations, explaining the *ARH* and *QRH* portfolio returns in the time series. Their raw return correlation is 98%. However, this correlation contains shared stock market variation. After accounting for the five Fama and French (2015) factors plus momentum using a linear "six-factor" model, the correlation between the residual *ARH* and residual *QRH* returns remains a very respectable 80%. The *QRH* stock choices capture more than just generic market or other known factor aspects.

The calculations are repeated for the Barber and Odean (2000) data in Appendix Table A.I. The association of the *QRH* investment weights with the *ARH* investment weights (of 76%) is almost the same as the association of the analogous *QBO* investment weights with the *ABO* investment weights (of 75%). Similarly, the rate of return correlation (net of the market) between *ARH* and *QRH* of 88% is almost the same as the (unreported net-of-the-market) correlation between *ABO* and *QBO* of 86%. And the (reported) correlations of return residuals net of the six-factor model are 82% for *ABO*–*QBO* versus 82% for *ARH*–*QRH*.

In sum, a good description of portfolios of retail favorites (relative to other known factor portfolios) is that they tilt heavily towards holdings of stocks that have had high trading volume over the last twelve months.

[20] The empirical coefficients are robust and not greatly sample-specific. (An earlier draft used out-of-sample weights and came to similar conclusions.) The two-thirds versus one-thirds combination predicted better than the 50-50 combination, but the two variables are sufficiently correlated that a 50-50 combination would have been acceptable too.

[21] The same portfolio weight normalization may have to be applied a second time to maintain a total investment weight of 100% when future returns can be missing.

V Holding Portfolio Return Performance

The *ARH* portfolio was investible in time—the holdings were downloadable from RT or from the RH API. Again, it is a crowd wisdom portfolio that should not be viewed as representative of the portfolios of individual investors. (Indeed, if reports are to be believed, with about 40 million holdings as of mid-2020 and given its base of about 13 million investors, the average RH investor only held about three positions, not 3,000!)

A. The Performance of the ARH *Crowd Consensus Portfolio*

[Insert Table VIII here]

Table VIII analyzes the predictive abnormal return performance of the *ARH* portfolio using the familiar methods of Fama and French (2015).

Panel A of Table VIII shows that the average daily performance of the *ARH* portfolio was a positive 10 bp per day using the zero-factor model, that is, net of the prevailing risk-free rate (from Morningstar via Ken French's popular data website). This abnormal performance declines to about 5 bp per day on the one-factor model, that is, when the (ex-post realized) market exposure of β=1.13 is taken into account. The abnormal performance increases back to 6.5 bp per day on the six-factor model. (Not shown, it was similarly positive for the five-factor model.) Not surprisingly, because the *ARH* portfolio moves slowly, its performance is insensitive to delay. With a five-day delay before investing, the performance deteriorates only a little, from an alpha of 10.4 bp per day to an alpha of 9.7 bp per day.

Panel B switches to monthly returns. This changes the rebalancing interval on both the factors and the *ARH* portfolio. The performance of the *ARH* portfolio is modestly better, offering an alpha of 1.3% per month against the six-factor model with a t-statistic of 2.46. Again, the portfolio is inert enough that even a three-month delay reduces this alpha only to 1.2% per month (with a t-statistic of 2.43).

Panel C calculates the rates of return on two portfolios, one formed at the end of June 2018 and the other at the end of June 2019, both held for one year.[22] The *ARH* portfolio seems to have performed well because it happened to avoid the negative returns of SMB, HML, and CMA. Statistical significance for two observations is of course impossible.

In sum, the *ARH* portfolio performed surprisingly well. Moreover, these Fama-French performance regressions assume equal investments in different periods. They ignore the effects of time-varying RH investor entry and exit into the stock market. The cross-sectional

[22]The factors are cumulated, because I did not have access to the long and short legs of the RMW and CMA portfolio for calculating mid-year compound rates of return. However, this should matter little.

performance here thus understates the performance due to RH investors having entered the stock market with good timing during a period of rising stock prices.

Nevertheless, caution is advisable. Although the alphas and their statistical significance are surprisingly high, investment performance—whether published in prestigious finance journals or not—is in general a notoriously poor predictor of future investment performance. It seems prudent to bear in mind that the solid finding here is not that RH investors can collectively beat the market, but rather that they were not cannon fodder for more sophisticated investors. This result can help explain why they did not attrition out quickly, but continued pouring in.

B. The Performance of the QRH *Proxy Portfolio*

[Insert Table IX here]

Table IX investigates the performance of the *QRH* proxy portfolio in the same 2018 to 2020 sample. (Recall that an equivalent *QBO* portfolio was also a good predictor of the HH-EW portfolio in the Barber and Odean (2000) data.) By its nature, *QRH*'s performance is likely to be similar to the performance of other volume-based trading strategies. The *QRH* performance in the 2018 to 2020 sample was about 4 bp per day lower than the *ARH* performance in all benchmark models. Over longer rebalancing intervals, the zero-factor and six-factor alphas remain positive, but the one-factor alpha turns negative.

Panel C explains the *ARH* portfolio performance including *QRH* as an additional "quasi-volume factor." Interestingly, the *QRH* volume factor can subsume most of the explanatory power of the more conventional Fama and French (2015) factors in explaining the *ARH* performance. The alpha reduction from 1.29% per month to 0.67% per month suggests that about half of *ARH*'s performance was related to trading volume, while the other half was due to unidentified skill or luck.[23] In sum, although trading volume was an important ingredient in RH holdings and contributed to the positive alpha on the zero-factor and six-factor models, the RH crowd knew better how to invest than with a naïve two-attribute *QRH* model.

[Insert Table X here]

Table X extends the *QRH*-based trading volume strategy back to 1980. Though investible (and/or also interpretable as merely the returns to a particular kind of volume-based

[23] It is possible to improve the predictive performance of *QRH* by taking the six-factor loadings into account. A modestly better proxy for the *ARH* investment weight would be $w_{\text{QRH}} \approx 0.15\times\text{HML}-0.15\times\text{UMD}-0.35\times\text{CMA}$. The 40-year inference in Table X remains similar under this measure.

strategy), these returns can also be viewed speculatively as standins to mimick the RH or Barber–Odean investors—akin to the first stage of an IV regression. The *QRH* portfolio would have achieved, by and large, similar performance over the extended 1980 to 2020 period as it did over the 2018 to 2020 period. It would have had positive zero-factor and six-factor alpha but negative one-factor alpha.

VI Trading Frictions

[Insert Table XI here]

Table XI shows that my data contain 9.06 million positions (i.e., investor stock holdings) matched to CRSP as of the end of 2019, 15.74 million positions as of the end of Q1 2020, and 26.52 million positions as of the end of Q2 2020. These observations mean that at least 6.64 and 10.78 million new positions would have had to have been established to account just for the inception of positions. The overnight holding changes in the *ARH* portfolio (i.e., the sums of absolute weight changes) over the full quarters were modestly larger, at 8.70 and 16.49 million positions.

It is likely that orders of less than 100 shares (probably typical for RH customers) would execute at or within the quoted bid-ask spread on the exchange, with only enough improvement not to raise the interest of the SEC. Boehmer, et al. (2020 even identify trades near, but not exactly equal to, bid and ask bounds as retail trades. My own personal experience at another broker similarly suggests that most retail market orders execute at the quoted bid-ask spread, which itself often stands at just 1 cent per share in the sample period.

I can assess trading costs further under one additional assumption: that the composition of *ARH* holdings and/or trading (absolute overnight changes) was representative of the trading of RH investors in general. This assumption extrapolates from observable holding and/or night-to-night trading patterns to unobservable trading patterns. It could be violated, for example, if certain types of customers liked to hold/trade certain stocks only during trading hours, or certain stocks traded hands only among RH investors (thus resulting in lower turnover—here, zero changes—in my data despite unusually active trading).

Under this assumption, Table XI suggests that an investor could have mimicked the crowd consensus portfolio with rather modest trading activity, above and beyond the need to establish the positions. The number of daily trades per quarter was about half as large as the total number of positions mid-quarter.

Not shown, using CRSP's quoted bid-ask spread as execution cost, the monthly rebalanced *ARH* buy-and-hold portfolio in Table VIII, with its rate of return of about 2%, would have suffered a drag of about 0.01% from rebalancing at the bid-ask spread (ignoring the original one-time incept cost). The daily portfolio would have traded only modestly more. In sum, transaction costs are of third-order importance to the monthly *ARH* portfolio rate of return.

[Insert Table XII here]

Table XII shows the proportional transaction costs of a daily *ARH* portfolio, winsorized at the 1% and 99% levels. Weighted by holdings, RH customers paid on average about 15 bp per dollar invested. The average spread for a $20 share would have been 3 cents. The trading-activity-weighted spread, which assumes buying or selling shares each night to maintain the *ARH* portfolio, would have been 2 cents.

The next set of estimates quote transaction drag in terms of the number of shares. These estimates allow for comparison with the public data in Table I, which suggests that RH itself earned about $0.24 per 100 shares in Q1 2020 and $0.17 per 100 shares in Q2 2020. The median half quoted bid-ask spread, expressed as a rate per 100 shares, was $0.50, which reflects a bid-ask spread of 1 cent. The more relevant mean quoted half-spreads were about $2.16 and $2.71 per 100 shares. Thus, RH received payment for order flow of about 10% of the mid-to-quote price. It is not known how the remaining 90% was split between price improvement to the RH customer, cost and profit to the market-maker, and sharing thereof with other parties.

RH investors were almost surely not efficient and frugal traders, maintaining the *ARH* portfolio with only the minimum required trades. Instead, millions of RH investors traded billions of shares. Unfortunately, it is difficult to assess whether this is because they held and traded positions of thousands of shares or whether they traded thousands of times in and out of one security each.[24]

Nevertheless, for many small equity investors, say, those with an account size and/or quarterly trading amount of $1,000, even the full bid-ask transaction cost scenario would likely amount only to about $1.50 per quarter. This is still only about 15% of the typical single-trade commission of $10 that most retail brokerage firms charged before the appearance of RH. In addition, small investors also received a free one-time sign-up stock allocation, typically worth about $3 to $5, thus covering about half a year's worth of trading.

[24] In the extreme, my calculations cannot detect (i) buying and selling between the polling intervals (one month in the above drag calculation, one day in most of my paper, and one hour in the original data) or (ii) mutual trading, where one RH investor buys and another sells. I also do not know what fraction of orders were limit orders versus market orders (Linnainmaa (2010)). This opaqueness is not unusual. No brokers release more information, and a federal statute (related to price manipulation) criminalizes the probing of execution quality by placing some round-trip trades. This law bizarrely ensures that researchers or even clients can never systematically experiment to learn their brokers' relative comparative transaction costs.

Finally, some RH customers probably view their trading as having some entertainment value. This is consistent with the unusually large presence of experience stocks in their portfolios. Although RH may or may not be the right retail broker for serious investors with large trades, it is a cheap thrill for exploring investors with small trades at a very low cost.

VII Conclusion

RH investors not only increased their relative holdings in individual stocks when stock prices increased or decreased greatly, but also increased their overall holdings during the market-wide 33% COVID drop in March 2020. Collectively, they did not panic. The fact that they added more positions four days after large market decreases (and large increases), suggests that they transferred cash to their RH accounts to fund more purchases. In March 2020, they were thus a (small) stabilizing force. Given the subsequent rise in the stock market, their timing and steadfastness contributed to their good portfolio returns—as did their general increase in participation from 2018 to 2020.

Some of the RH investors' holdings seem bizarre. For example, they fell in love with some obscure experience stocks, such as cannabis stocks. Nevertheless, on the whole, the *ARH* crowd consensus portfolio (itself likely similar to a household-equal-weighted portfolio) was not greatly tilted toward these experience stocks. Instead, a better description is that RH investors tilted mostly towards stocks with above-average trading volume over the previous 12 months. Remarkably, this association is strong enough to allow the use of a trading volume–derived factor as a proxy for the performance of stock holdings preferentially held by retail investors, at least in some circumstances. Visual inspection further suggests that RH investors had a preference for stocks of firms with products that they were familiar with. As of August 2000, the end of the data, they had turned towards familiar but fallen old-economy stalwarts like airline companies.

From the mid-2018s to the mid-2020s, the RH consensus portfolio performed well in the cross-section, earning positive alphas with respect to the risk-free rate, the market model, and a Fama and French (2015) five-factor plus momentum model. Past performance is no guarantee of future performance, but good performance can explain why RH investors in aggregate have not attritioned out but continued to pour in.

References

Barber, Brad M., Xing Huang, Terrance Odean, and Chris Schwarz, 2020, Attention induced trading and returns: Evidence from Robinhood users, Technical report, University of California at Berkeley.

Barber, Brad M., and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *Journal of Finance* 55, 773–806.

Barber, Brad M., and Terrance Odean, 2002, Online investors: Do the slow die first?, *Review of Financial Studies* 15, 455–488.

Barber, Brad M., and Terrance Odean, 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.

Barber, Brad M., and Terrance Odean, 2013, The behavior of individual investors, in George M. Constantinides, Milton Harris, and Rene M. Stulz, eds., *Handbook of the Economics of Finance*, chapter 22, 1535–1565 (Elsevier).

Barber, Brad M., Terrance Odean, and Ning Zhu, 2008, Do retail trades move markets?, *Review of Financial Studies* 22, 151–186.

Ben-David, Itzhak, Francesco Franzoni, Byungwook Kim, and Rabih Moussawi, 2020, Competition for attention in the ETF space, Technical report, Ohio State University, Lugano, Swiss Finance Institute, and Villanova.

Ben-Rephael, Azi, Zhi Da, and Ryan D. Israelsen, 2017, It depends on where you search: Institutional investor attention and underreaction to news, *Review of Financial Studies* 30, 3009–3047.

Boehmer, Ekkehart, Charles M. Jones, Xiaoyan Zhang, and Xinran Zhang, 2021, Tracking retail investor activity, *Journal of Finance, Forthcoming* TBA, TBA.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chen, Joseph, Harrison Hong, and Jeremy C. Stein, 2002, Breadth of ownership and stock returns, *Journal of Financial Economics* 66, 171–205.

Cheng, Peter, Thomas J. Murphy, and Marko Kolanovic, 2020, Follow the robinhood money, Technical report, J.P. Morgan.

Choi, James J., Li Jin, and Hongjun Yan, 2013, What does stock ownership breadth measure?, *Review of Finance* 17, 1239–1278.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011, In search of attention, *Journal of Finance* 66, 1461–1499.

DellaVigna, Stefano, and Joshua M. Pollet, 2009, Investor inattention and friday earnings announcements, *Journal of Finance* 64, 709–749.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fang, Lily, and Joel Peress, 2009, Media coverage and the cross-section of stock returns, *Journal of Finance* 64, 2023–2052.

Fong, Kingsley Y.L., David R. Gallagher, and Adrian D. Lee, 2014, Individual investors and broker types, *Journal of Financial and Quantitative Analysis* 49, 431–451.

Glossner, Simon, Pedro Matos, Stefano Ramelli, and Alexander F. Wagner, 2020, Where do institutional investors seek shelter when disaster strikes? Evidence from COVID-19, Technical report, Virginia/Darden and Zurich.

Grinblatt, Mark, and Matti 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.

Grinblatt, Mark, and Matti Keloharju, 2001, What makes investors trade?, *Journal of Finance* 56, 589–616.

Harms, Torsten, 2021, Generation stupid or generation smart? an analysis of young investors in germany and austria, Technical report, Karlsruhe.

Hirshleifer, David, Sonya Seongyeon Lim, and Siew Hong Teoh, 2009, Driven to distraction: Extraneous events and underreaction to earnings news, *Journal of Finance* 64, 2289–2325.

Kaniel, Ron, Shuming Liu, Gideon Saar, and Sheridan Titman, 2012, Individual investor trading and return patterns around earnings announcements, *Journal of Finance* 67, 639–680.

Kelley, Eric K., and Paul C. Tetlock, 2013, How wise are crowds? Insights from retail orders and stock returns, *Journal of Finance* 68, 1229–1265.

Kumar, Alok, 2009, Who gambles in the stock market?, *Journal of Finance* 64, 1889–1933.

Lee, Charles M.C., and Mark J. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733–746.

Linnainmaa, Juhani T., 2010, Do limit orders alter inferences about investor performance and behavior?, *Journal of Finance* 65, 1473–1506.

Moss, Austin, James P. Naughton, and Clare Wang, 2020, The irrelevance of ESG disclosure to retail investors: Evidence from Robinhood, Technical report, University of Iowa, University of Virgina, and University of Colorado.

Nagel, Stefan, 2005, Short sales, institutional investors and the cross-section of stock returns, *Journal of Financial Economics* 78, 277–309.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Ozik, Gideon, Ronnie Sadka, and Siyi Chen, 2020, Flattening the illiquidity curve: Retail trading during the COVID-19 lockdown, Technical report, Edhec and Boston College and Chinese University of Hong Kong.

Peng, Lin, and Wei Xiong, 2006, Investor attention, overconfidence and category learning, *Journal of Financial Economics* 80, 563–602.

Seru, Amit, Tyler Shumway, and Noah Stoffman, 2010, Learning by trading, *Review of Financial Studies* 23, 705–739.

Zarnowitz, Victor, and Louis A. Lambros, 1987, Consensus and uncertainty in economic prediction, *Journal of Political Economy* 95, 591–621.

Table I

Payment for Order Flow and Implied Trading Volume

The data for the first four rows come from a table from Piper-Sandler as reported on CNBC on 08/13/2020. The last two rows come from NASDAQ's 2020/2021 reported quarterly trading volumes and market shares. "/100" indicates per 100 shares traded. "Implied #" indicates the number of shares traded imputed from reported total payment and payment rate. The total U.S. equities estimate is calculated from NASDAQ's statement of its market share for U.S. equities. Payments for NASDAQ are reported revenues.

	Q1 2020			Q2 2020		
	Payment	Rate/100	Implied #	Payment	Rate/100	Implied #
Robinhood	\$31m	\$0.24	13 billion	\$69m	\$0.17	41 billion
Schwab	\$25m	\$0.11	23 billion	\$32m	\$0.11	29 billion
ETrade	\$30m	\$0.16	19 billion	\$50m	\$0.15	33 billion
TD Ameritrade	\$73m	\$0.15	49 billion	\$144m	\$0.15	96 billion
Sum			104 billion			199 billion
NASDAQ U.S. Equities	\$63m	\$0.050	127 billion	\$74m	\$0.052	143 billion
Total U.S. Equities			678 billion			780 billion

Table II

Extreme One-Day Increases in RH Holdings

The table reports the 10 cases with the largest one-day *ARH* investment weight increases. The $\Delta_{-1,0}\#$ column gives the exact number of RH holding changes (subscripts refer to trading days). The preceding columns round to thousands. The weight w_0 is the weight of stock i on day 0, as in the *ARH* portfolio, $w_{i,t} \equiv N_{i,t} / \sum_i N_{i,t}$, where N_i is the number of RH investors in stock i on day t. The sort column is $\Delta_{-1,0}w$. *Interpretation:* There are typically large positive or negative daily rates of return before or around large increases in the number of RH investors.

Panel A: RH Holdings in Thousands

	Date	Ticker	–2	–1	±0	+1	+2	$\Delta_{-1,0}\#$
1	07/26/2018	Facebook	108	114	156	166	170	42,083
2	01/16/2020	Inpixon	0	0	26	26	25	26,376
3	10/02/2018	Oragenics	17	27	41	39	37	14,336
4	07/17/2018	Netflix	101	107	118	118	117	11,000
5	11/01/2019	Fitbit	252	252	271	275	275	18,924
6	03/06/2020	Inovio Pharm	107	111	137	155	150	25,521
7	06/03/2020	Genius Brands	49	66	112	145	141	46,014
8	07/27/2018	Facebook	114	156	166	170	173	9,832
9	06/09/2020	Nikola	33	77	125	130	137	48,019
10	05/18/2020	Sorrenta Thera	16	51	92	92	90	41,302

Panel B: RH Change and Stock Rate of Return

		ARH		Rate of Return			
	Ticker	w_0	$\Delta_{-1,0}w$	r_{-1}	r_0	r_{+1}	Possible Reason
1	FB	3.22	0.85	0.4	–18.7	–0.1	Q2 earnings below exp
2	INPX	0.27	0.27	–55.5	8.1	–7.2	1/8: Rev split, some good.
3	OGEN	0.74	0.25	38.8	123.2	–51.1	Various minor news
4	NFLX	2.49	0.23	1.3	–5.6	–1.4	Poor earnings, Walmart
5	FIT	3.06	0.21	5.8	14.6	–1.8	Google buys Fitbit
6	INO	1.16	0.20	25.5	45.5	–22.6	Accelerated COVID vaccine
7	GNUS	0.46	0.19	52.6	95.9	–13.2	Funding for children network
8	FB	3.40	0.19	–18.7	–0.1	–1.6	See #1; acquires Redkix
9	NKLA	0.49	0.18	102.5	9.6	–17.9	Progress. See Table III#3
10	SRNE	0.41	0.18	157.6	–7.0	–15.6	Positive Covid treatment news

Table III

Extreme One-Day Decreases in RH Holdings

This table lists the 10 cases with the largest one-day *ARH* investment weight decreases. For more explanations, refer to Table II. *Interpretation:* There are sometimes large positive or negative daily rates of return around large decreases in *ARH* weight. The relation is weaker than that in Table II.

Panel A: RH Holdings in Thousands

	Date	Ticker	–2	–1	±0	+1	+2	$\Delta_{-1,0}\#$
–1	02/27/2019	India Global.	78	78	42	42	42	–36,084
–2	11/02/2018	Inpixon	39	40	1	1	1	–38,583
–3	06/04/2020	Nikola	95	100	21	33	77	–78,685
–4	02/05/2020	Telsa	154	163	148	151	152	–15,339
–5	03/13/2020	Aikido Pharm	18	18	1	1	1	–17,240
–6	02/04/2019	Ohr Pharm	8	8	0	0	0	–8,742
–7	10/01/2018	Tesla	95	101	94	94	94	–6,876
–8	10/04/2018	India Global	89	83	76	72	80	–7,259
–9	01/16/2020	Fitbit	257	257	256	257	259	–898
–10	08/02/2018	Tesla	85	85	80	77	76	–4,857

Panel B: RH Change and Stock Rate of Return

		ARH		Rate of Return			
	Ticker	w_0	$\Delta_{-1,0}w$	r_{-1}	r_0	r_{+1}	Possible Reason
–1	IGC	0.63	–0.71	–2.1	1.9	13.7	Relists on NYSE
–2	INPX	0.03	–0.67	–3.5	–31.5	1.8	Reverse stock split
–3	NKLA	0.09	–0.32	6.9	–0.3	4.0	Various. see also #9
–4	TSLA	1.43	–0.15	12.2	–18.3	1.6	Good earnings; mixed other news
–5	AIKI	0.01	–0.14	–8.6	–33.8	–12.0	Name change
–6	OHRP	0.00	–0.14	–4.6	–23.7	–2.5	Reverse stock split
–7	TSLA	1.69	–0.13	–13.9	17.0	–3.1	Musk settles with SEC
–8	IGC	1.35	–0.13	–32.0	–26.8	–36.3	10/2: ATM offering completed
–9	FIT	2.59	–0.11	0.6	–0.5	1.0	Scripps: Fitbit may detect flu
–10	TSLA	1.61	–0.11	1.0	15.7	–0.9	OK earnings, China plan

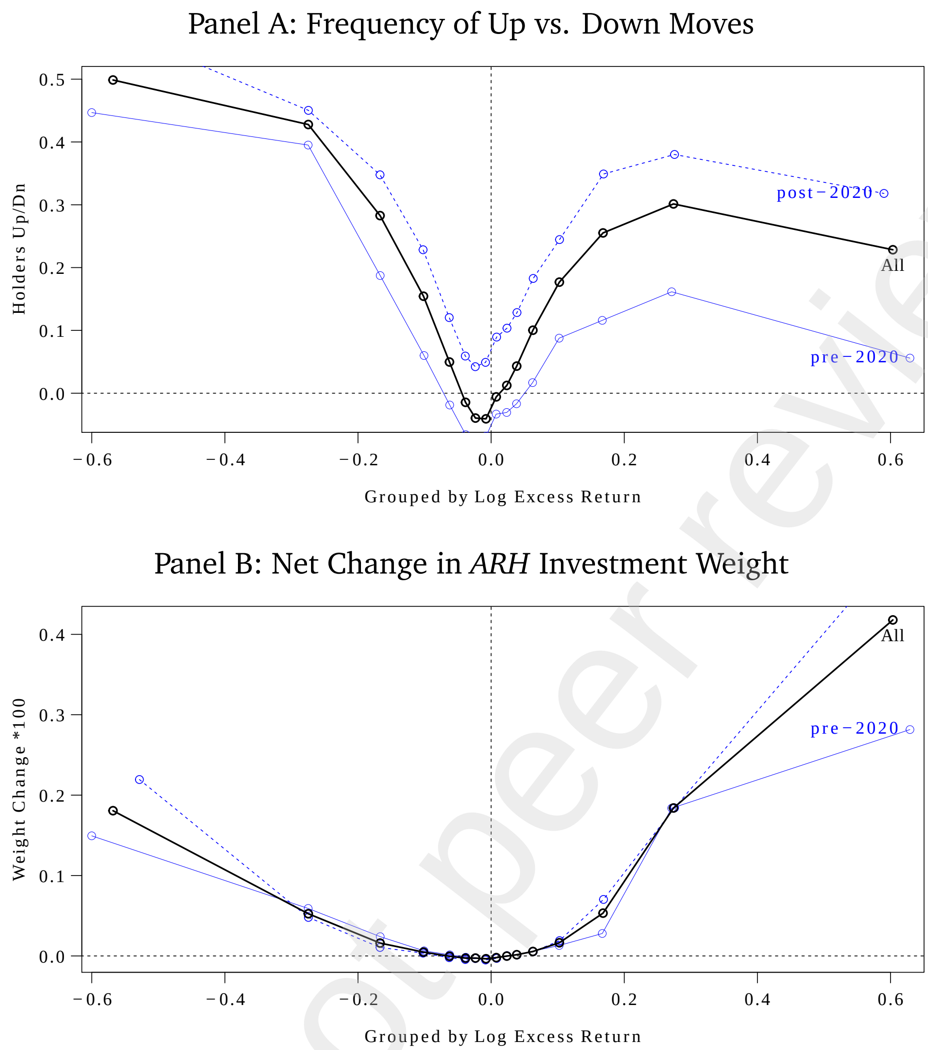

Figure 1. RH holding changes by previous day net-of-market rate of return. Stock–days are first grouped by net-of-market stock returns into about 20 (not equally spaced) categories. Within each category, the y-axis presents mean (change) statistics for the full sample (dark line) and two subsamples (light lines, pre-2020 and post-2020) for the subsequent day. Panel A assigns +1 and –1 to stock–days on which the RH number increased and decreased, respectively, and plots the mean over all stock–days in each rate of return bin. Panel B shows the net change in the *ARH* investment weight. This weights a larger number of investor changes more (common in large stocks with more RH holders) and also considers other *ARH* changes on the same day. The sample begins in June 2018 and ends in September 2020. *Interpretation:* RH investors preferentially purchase large movers. This is similar to the behavior of other retail investors described in earlier work. Contrarian increases are concentrated in small stocks with smaller weight increases relative to the increases in the total number of RH holdings.

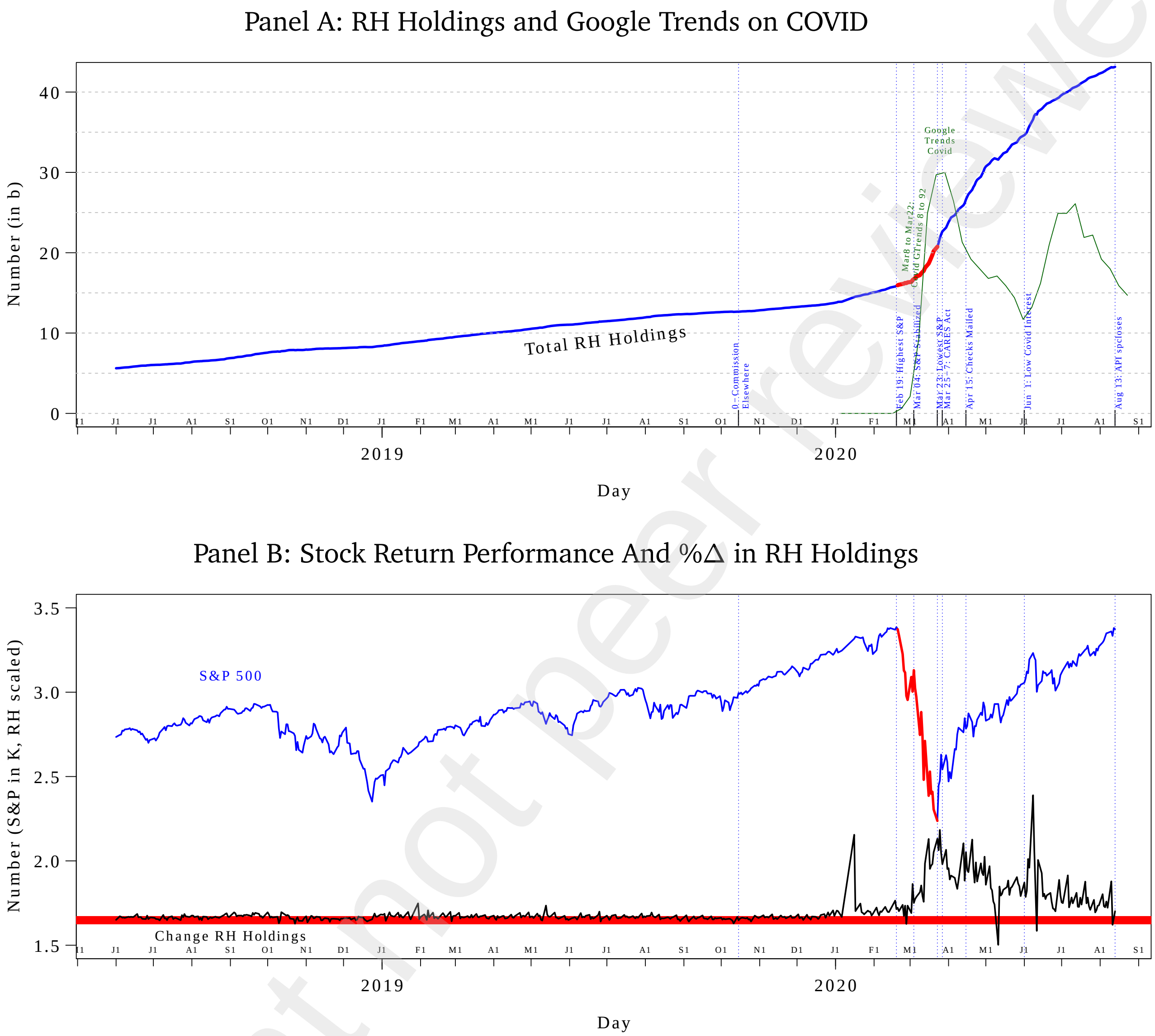

Figure 2. Sum-total RH holdings and stock market performance, 2018 to 2020. Panel A shows the total number of RH investors. The green line shows the public interest in COVID according to Google Trends. Panel B shows the S&P 500 index and the percentage change in the sum-total of RH holdings. The red line indicates zero growth. (Figure A1 in the Appendix is a closeup for 2020.) *Interpretation:* RH investing accelerated during the U.S. COVID crisis.

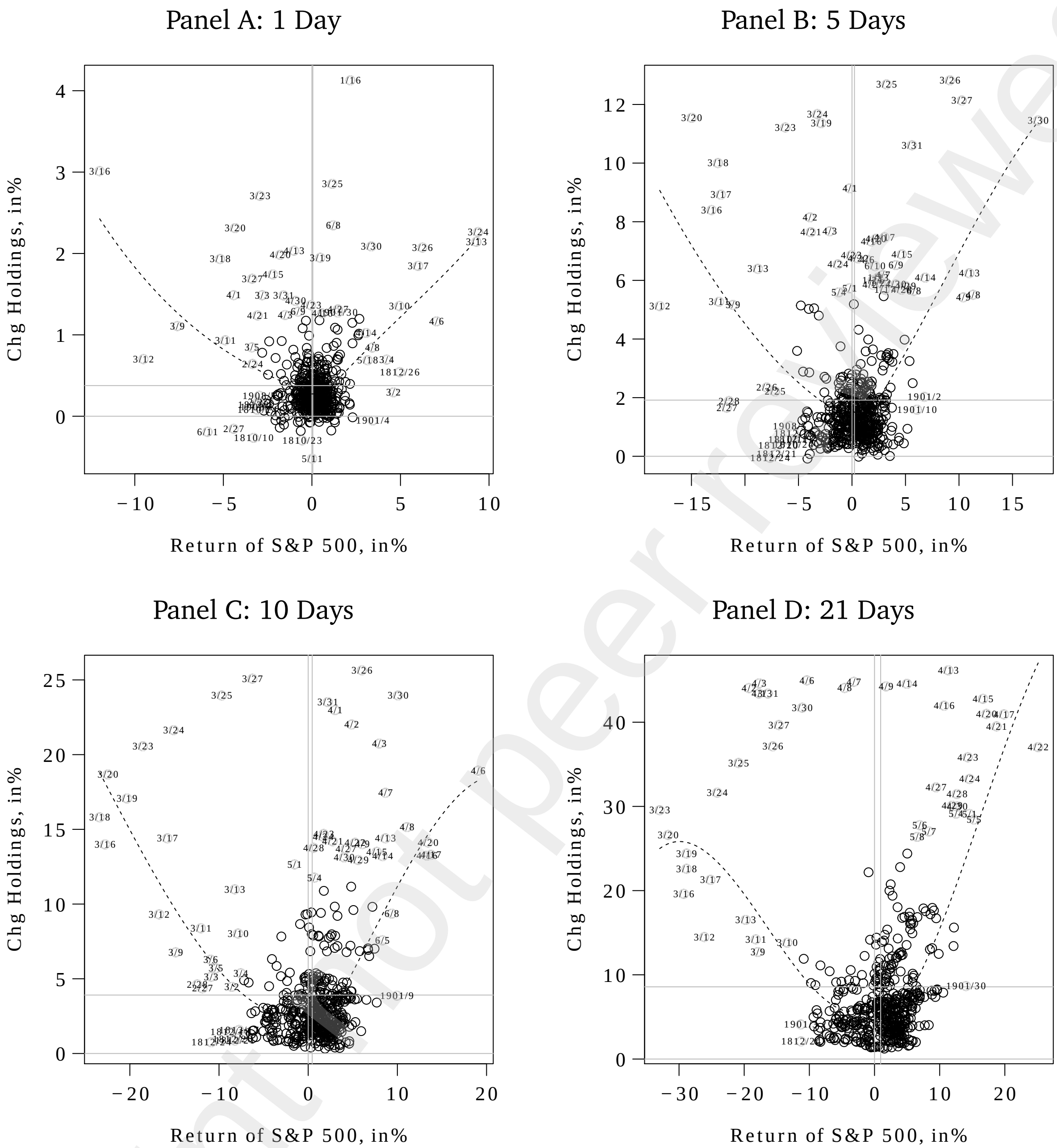

Figure 3. 2018 to 2020 RH systematic contrarianism by horizon. This figure shows the percentage change in the sum-total of RH holdings versus the S&P 500 rate of return, similar to Table II, but in *x*-*y* rather than time-series format. The aggregation periods are contemporaneous, equally long for *x* and *y* variables, and overlapping for the longer intervals. Days with large changes are indicated. The lines are fitted by a local polynomial regression (R loess) with a span of 0.75. *Interpretation:* Large S&P 500 price drops did not deter RH investors.

Table IV

Daily Percent Changes in the Growth Rate of RH Holdings

The dependent variable (y) in this time-series regression is the percentage change in the sum-total number of RH holdings, $Y_t \equiv \%\Delta_t \sum_i \text{RH}_{it}$. The market rate of return is the percentage change in the S&P 500. The "positive market movement" variables assign zero to the variable on days when the market declined. The "negative market movement" variables are analogous, except they use the absolute value of R_M, as indicated in the header, to simplify interpretation. (A positive coefficient means an increase in holdings.) t-statistics are adjusted for heteroskedasticity and two lags as in Newey and West (1987) (without prewhitening). *Interpretation:* Zero to one-day and four to five-day-lagged large market movements are robust positive predictors of increases in the sum-total number of RH holdings.

	06/01/2018 to 08/13/2020				02/03/2020 to 08/13/2020			
	Coef	T	Coef	T	Coef	T	Coef	T
(Intercept)	0.08	2.60	0.03	1.84	0.07	1.77	0.03	0.65
Lagged Values								
lag(Y, 1)	0.52	4.95	0.24	2.69	0.39	4.39	0.30	3.49
lag(Y, 2)			0.26	5.06	0.19	1.84	0.17	1.75
lag(Y, 3)			0.21	3.15	0.16	1.81	–0.07	–0.70
lag(Y, 4)							0.20	2.38
lag(Y, 5)							0.06	0.72
lag(Y, 6)							–0.12	–1.58
Positive Market Movement: $(R_M>0)\cdot R_M$								
lagp(R_M, 0)							0.07	2.48
lagp(R_M, 1)	0.13	5.20	0.09	3.09	0.11	2.54	0.13	3.92
lagp(R_M, 2)			0.00	–0.14	–0.03	–0.83	–0.04	–1.47
lagp(R_M, 3)			–0.01	–0.59	–0.02	–0.83	–0.03	–1.14
lagp(R_M, 4)							0.12	4.14
lagp(R_M, 5)							0.03	1.23
lagp(R_M, 6)							0.01	0.19
Negative Market Movement, Absolute Value: $(R_M<0)\cdot\lvert R_M\rvert$								
lagn(R_M, 0)							0.00	0.17
lagn(R_M, 1)	0.09	3.99	0.06	3.35	0.08	2.81	0.04	2.13
lagn(R_M, 2)			0.02	1.08	0.01	0.43	–0.02	–0.84
lagn(R_M, 3)			0.00	0.11	–0.00	–0.01	–0.02	–0.89
lagn(R_M, 4)							0.05	2.19
lagn(R_M, 5)							0.02	0.82
lagn(R_M, 6)							–0.01	–0.30
$\bar{R}^2$	51.6%		59.4%		67.6%		73.4%	
N	543		541		132		129	

Table V

Highest *ARH* vs. VWMKT Log-Investment Rank Differences on 12/31/2019 or 06/30/2020

RH refers to Robinhood and VW to the value-weighted market (according to CRSP's number of shares times price). The stocks listed in this table (log-)ranked more highly in the investment holdings of RH investors than in the value-weighted market portfolio on 12/31/2019 or on 06/30/2020, such that they constituted at least 0.15% of the *ARH* investment weight. (All stocks listed in this table constituted less than 0.005% of the holdings in the VW portfolio.)

		Business	Rank		*ARH* Pfio
Year	TIC	Description	RH	VW	Weight
2019	BIOC	Cancer-Detection	102	3356	**0.15**
2019	HUSA	Oil-Gas	88	3425	**0.19**
2019	IGC	India-Cannabis	27	3271	**0.59**
2019	OGEN	BioPharm Immunization	39	3280	**0.47**
2019	RIOT	Blockchain Financial	96	3161	**0.17**
2019	ZN	Oil-Gas	95	3296	**0.17**
2020	CEI	Oil-Gas	94	3425	**0.20**
2020	INPX	Big Data	88	3227	**0.22**
2020	OCGN	Eye Disease	101	3403	**0.19**
2020	TTNP	Drug Implants	90	3274	**0.22**

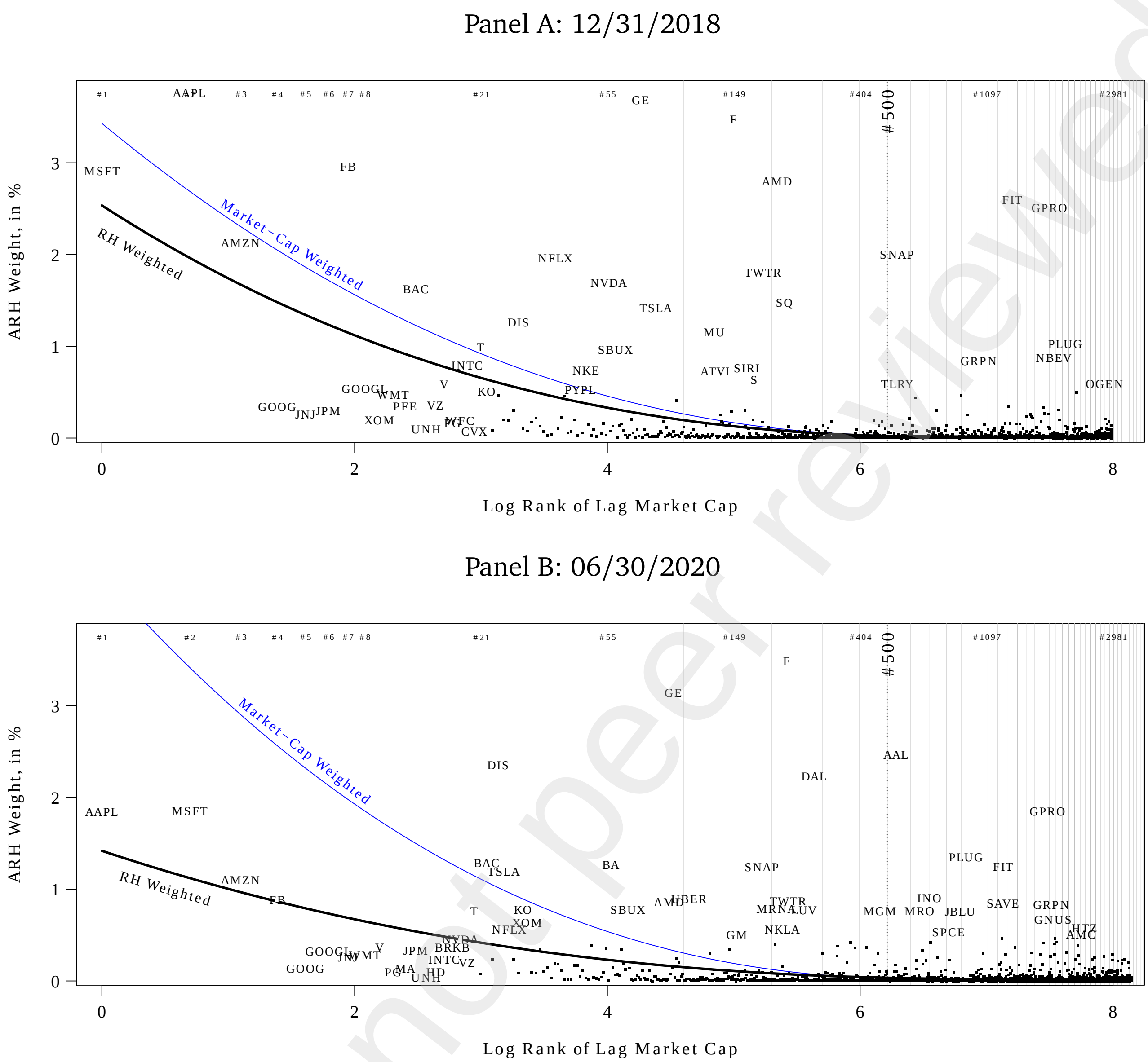

Figure 4. Snapshot *ARH* weights by log-rank of market cap. Each point is one stock's percentage investment weight in *ARH*. These weights are plotted against the log market-cap rank (from CRSP). For example, although AAPL was the biggest stock, accounting for 5.2% of the total CRSP market cap on 06/30/2020, it accounted for "only" 1.85% of the *ARH* portfolio (as marked). For stocks with unusually large *ARH* investment weights, and for the biggest stocks, the plots show ticker symbols instead of points. The black line fits the (RH) points. For comparison, the blue line shows a smoothed line that a market-cap weighted portfolio would have assigned. Both lines are fitted via R loess with a span of 0.75. *Interpretation:* Relative to a value-weighted portfolio, RH investors typically underinvested in the biggest 500 stocks, especially in mid-2020. They overinvested in many consumer-related tech stocks, as well as fallen angels (such as Ford (F), General Electric (GE), and United Airlines (UAL)).

Table VI

***ARH* Investment Weights *Far Above* VWMKT Investment Weights**

Listed are the companies with the largest (absolute) excess *ARH* weights over the value-weighted market cap (VWM) weights on 12/31/2018 or 06/30/2020. The holdings changes and rates of return in the last four columns are from 12/31/2018 to 06/30/2020. *Interpretation:* Some companies increased their *ARH* minus VW weights through active RH purchases, some due to less purchasing of other stocks, and some due to price drops affecting the VW weight.

Panel A: Rank and Weights

		Pre-COVID				Post-COVID			
		Dec 2018				Jun 2020			
		Rank		Weight		Rank		Weight	
Tic	Description	*ARH*	VW	*ARH*	VW	*ARH*	VW	*ARH*	VW
AAL	American Airlines	114	272	0.12	0.06	3	620	2.39	0.02
AMD	AMD	6	218	2.70	0.08	21	92	0.84	0.20
BA	Boeing	53	26	0.29	0.78	11	57	1.23	0.34
DAL	Delta Airlines	82	139	0.17	0.15	5	288	2.16	0.06
DIS	Disney	17	27	1.22	0.69	4	23	2.28	0.66
F	Ford	3	152	3.35	0.13	1	233	3.38	0.08
FB	Facebook	4	7	2.85	1.34	19	4	0.86	1.78
FIT	Fitbit	7	1388	2.51	0.00	13	1281	1.21	0.01
GE	General Electric	2	72	3.55	0.28	2	95	3.05	0.20
GPRO	GoPro	8	1867	2.42	0.00	7	1785	1.79	0.00
INO	Vaccines	384	1931	0.03	0.00	17	725	0.88	0.01
JBLU	JetBlue Airlines	100	587	0.14	0.02	31	920	0.72	0.01
LUV	Southwest Airlines	51	167	0.29	0.11	27	266	0.75	0.07
MGM	MGM Casino-Entmt	180	301	0.07	0.05	30	488	0.74	0.03
MRO	Marathon Oil	129	322	0.10	0.05	29	668	0.74	0.02
MU	Micron Storage	18	131	1.11	0.15	97	99	0.19	0.19
NBEV	Cannabis Drinks	24	1927	0.84	0.00	87	2637	0.23	0.00
NFLX	Netflix	11	37	1.89	0.50	36	25	0.54	0.65
NVDA	Nvidia	13	56	1.63	0.35	42	17	0.44	0.76
PLUG	Hydrogen Fuel	19	2115	0.99	0.00	9	971	1.31	0.01
SAVE	Spirit Airlines	501	689	0.02	0.02	22	1282	0.82	0.01
SNAP	Snapchat	10	559	1.93	0.02	14	189	1.21	0.09
SQ	Square Pymnts	15	230	1.41	0.07	61	145	0.33	0.12
TSLA	Tesla	16	82	1.37	0.24	15	24	1.16	0.65
TWTR	Twitter	12	194	1.74	0.09	20	236	0.84	0.08
UAL	United Airlines	335	188	0.04	0.10	12	437	1.23	0.03
XXII	Cannabis-Nctn	23	2050	0.85	0.00	103	2793	0.19	0.00
ZNGA	Gaming (FB)	21	757	0.94	0.01	23	464	0.77	0.03

Panel B: Holdings and Price Changes				
	Holdings (in k)			Price Chg
Tic	2018	2020	Chg	(RoR)
AAL	7	654	88.7	–0.64
AMD	161	229	0.4	1.92
BA	17	337	18.5	–0.42
DAL	10	592	56.1	–0.47
DIS	72	624	7.6	0.04
F	200	925	3.6	–0.18
FB	170	235	0.4	0.67
FIT	149	331	1.2	0.30
GE	212	834	2.9	–0.05
GPRO	144	491	2.4	0.19
INO	1	241	133.4	5.41
JBLU	8	197	23.0	–0.36
LUV	17	205	10.8	–0.33
MGM	4	202	44.3	–0.31
MRO	6	202	31.6	–0.55
MU	66	53	–0.2	0.42
NBEV	50	61	0.2	–0.76
NFLX	113	148	0.3	0.65
NVDA	97	121	0.2	2.13
PLUG	59	358	5.1	4.85
SAVE	1	225	174.0	–0.69
SNAP	115	330	1.9	2.66
SQ	84	89	0.1	0.68
TSLA	81	317	2.9	1.73
TWTR	104	231	1.2	–0.03
UAL	2	336	157.2	–0.60
XXII	51	51	0.0	–0.67
ZNGA	55	210	2.8	1.14
All *ARH*	5,973	27,390	3.6	0.24

Table VII

Correlations of *QRH* with *ARH*

The *QRH* portfolio is two-thirds of a one-year volume-weighted portfolio (*VOL*) and one-third of a one-year dollar volume-weighted portfolio (*DOLVOL*). These two variables and weights were roughly chosen based on an empirical exploration of many variables predicting *ARH* investment weights. The two *QRH* input variables are themselves correlated, and reasonable alternative weightings do not greatly change the return correlations. *Interpretation:* The *QRH* portfolio is an (albeit imperfect) proxy for the *ARH* portfolio.

Panel A: Correlations of Portfolio Investment Weights

		w(*VOL*)	w(*DOLVOL*)	w(*QRH*)
w(*ARH*)	on 12/31/2018	0.73	0.67	0.81
w(*ARH*)	on 12/31/2019	0.67	0.54	0.71
w(*ARH*)	Pooled XS and TS	0.71	0.43	0.76

Panel B: Correlations of Daily Portfolio Returns in the Time Series

Portfolio	Return Adjustment	Correlation
QRH	Raw	0.98
	Net of 1-F Model	0.87
	Net of 6-F Model	0.78
–20190631	Net of 6-F Model	0.80
20190631–	⋮	0.79
QRH	Net of 1×Market	0.88
SMB	⋮	0.05
HML	⋮	–0.05
RMW	⋮	0.25
CMA	⋮	–0.29
UMD	⋮	–0.48

Table VIII

Return Performance of the *ARH* Crowd Portfolio, in %

The *ARH* portfolio weights are based on the number of RH investors in each stock, as in equation (1) ($w_i \equiv N_i / \sum N_i$). The table shows the (net of risk-free) return performance of the *ARH* portfolio with respect to various rebalancing intervals, delays, and benchmarks. The 0-F benchmark is the mean (net of the prevailing risk-free rate). The 1-F benchmark is the CAPM. The 6-F benchmark is the Fama and French (2015) five-factor model plus UMD momentum. t-statistics are Newey and West (1987) adjusted (one lag, no pre-whitening). *Interpretation:* The *ARH* crowd portfolio did not underperform. Rebalanced daily, it tilted towards small (SMB), aggressive (CMA), past-loser (UMD) stocks.

Panel A. Daily Rebalancing, No Delay			
	0-F	1-F	6-F
alpha	0.104	0.050	0.065
(T)	(1.27)	(1.37)	(2.70)
XMKT		1.13	1.03
SMB			0.45
HML			–0.10
RMW			–0.21
CMA			–0.57
UMD			–0.35
546 Days (06/01/2018 – 08/14/2020)			

Panel B. Daily Rebalancing, 5-Day Delay			
	0-F	1-F	6-F
alpha	0.097	0.045	0.062
(T)	(1.19)	(1.28)	(2.63)
XMKT		1.13	1.04
SMB			0.44
HML			–0.11
RMW			–0.22
CMA			–0.55
UMD			–0.34
546 Days (06/07/2018 – 08/20/2020)			

Panel C. Monthly Rebalancing, No Delay			
	0-F	1-F	6-F
alpha	2.09	0.61	1.29
(T)	(1.46)	(1.13)	(2.46)
XMKT		1.30	1.09
SMB			0.89
HML			–0.36
RMW			–0.47
CMA			0.09
UMD			–0.19
27 months (06/2018 – 08/2020)			

Panel D. Monthly Rebalancing, 3-Mo Delay			
	0-F	1-F	6-F
alpha	1.92	0.75	1.20
(T)	(1.20)	(1.38)	(2.43)
XMKT		1.30	1.13
SMB			0.74
HML			–0.31
RMW			–0.30
CMA			0.04
UMD			–0.15
26 months (08/2018 – 09/2020)			

Panel E. Annual Mid-Year Rebalancing / Returns (Not Regression Coefficients)								
	ARH	XMKT	SMB	HML	RMW	CMA	UMD	RF
07/2018 - 06/2019	8.2	8.0	–13.1	–11.8	4.9	–1.5	–0.5	2.2
07/2019 - 06/2020	26.5	8.8	–9.5	–33.2	0.5	–5.6	7.1	1.3

Table IX

Return Performance of the *QRH* Proxy Portfolio, in %

Panels A and B are analogous to Table VIII with *QRH* excess return as the dependent variable instead of *ARH* excess return. *QRH* is two-thirds a number-of-shares trading volume variable (mapped into a weight as in $w_i(v_i) \equiv v_i / \sum_i v_i$) and one-third a dollar trading volume variable ($w_i(d_i) \equiv d_i / \sum_i d_i$). Panel C explores the *QRH* portfolio return as an independent factor explaining the *ARH* rate of return. *Interpretation: QRH* is not as good a predictor of future returns as *ARH*. The volume component in *QRH* can account for about half of the *ARH* alpha.

Panel A. Daily Rebalancing, No Delay

	0-F	1-F	6-F
alpha	0.059	0.005	0.029
(T)	(0.79)	(0.19)	(2.00)
XMKT		1.14	1.05
SMB			0.35
HML			0.06
RMW			–0.20
CMA			–0.24
UMD			–0.21
546 days (06/01/2018 - 08/14/2020)			

Panel B. Daily Rebalancing, 5-Day Delay

	0-F	1-F	6-F
alpha	0.052	0.000	0.027
(T)	(0.69)	(0.02)	(1.87)
XMKT		1.14	1.06
SMB			0.35
HML			0.06
RMW			–0.20
CMA			–0.24
UMD			–0.21
546 days (06/07/2018 - 08/20/2020)			

Panel C. Monthly Rebalancing, No Delay

	0-F	1-F	6-F
alpha	1.08	–0.45	0.40
(T)	(0.74)	(–1.65)	(1.37)
XMKT		1.35	1.21
SMB			0.56
HML			–0.05
RMW			–0.33
CMA			0.26
UMD			–0.10
27 Months (06/2018 to 08/2020)			

Panel D. Monthly Rebalancing, 3-Mo Delay

	0-F	1-F	6-F
alpha	0.86	–0.35	0.36
(T)	(0.56)	(–1.45)	(1.52)
XMKT		1.34	1.21
SMB			0.49
HML			–0.04
RMW			–0.24
CMA			0.19
UMD			–0.11
26 Months (08/2018 to 09/2020)			

Panel E. Including a *QRH* Volume Factor Component in *ARH* Return Regression, 27 months

	alpha	(T)	XMKT	SMB	HML	RMW	CMA	UMD	*QRH*
(I)	1.29	(2.46)	1.09	0.89	–0.36	–0.47	0.09	–0.19	
(II)	0.67	(1.74)	–0.24	0.27	–0.32	–0.08	–0.18	–0.07	1.11

Table X

Extended Return Performance of the *QRH* Proxy Portfolio After 1980, in %

The dependent variable is the rate of return on the *QRH* portfolio net of the risk-free rate (as used and described in the two previous tables). *QRH* is two-thirds the one-year volume-trading and one-third the one-year dollar-volume trading portfolio. Annual returns are for calendar years, with weights chosen in November of the preceding year. *Interpretation:* Over a 40-year interval, the *QRH* portfolio performed about the same as in the 2018 to 2020 sample.

Panel A. No Delay, Monthly Returns			
	0-F	1-F	6-F
alpha	0.60	–0.32	0.16
(T)	(1.96)	(–2.99)	(1.86)
XMKT		1.34	1.17
SMB			0.33
HML			–0.10
RMW			–0.44
CMA			–0.20
UMD			–0.24
489 months (01/1980 - 09/2020)			

Panel B. November, Annual Returns			
	0-F	1-F	6-F
alpha	8.0	–3.1	0.3
(T)	(2.59)	(–2.90)	(0.40)
XMKT		1.26	1.10
SMB			0.47
HML			–0.13
RMW			–0.44
CMA			0.00
40 years (1980 to 2019)			

Table XI

End-of-Quarter Holdings and Intraquarter Day-to-Day Changes

RH holdings are in million investors. IQ are intra-quarter changes based on the sum of all absolute overnight holding changes during the quarter (only for CRSP matched holdings) and thus reflect a daily rebalancing investment strategy. *Interpretation:* Much of the overnight changes could be accounted for by the growth in positions.

	12/31/2019		03/31/2020		06/30/2020
RH Holdings	13.75		23.46		39.51
CRSP-Matched Holdings	9.06		15.74		26.52
Delta		6.64		10.78	
RH IQ-Summed Day-to-Day Changes		8.70		16.49	

Table XII

Estimated Trading Frictions

This table shows weighted half-quoted spreads and share prices. Bid-ask quotes are end-of-day and from CRSP. RH Holding-Weighted means based on end-of-quarter holdings. IQ indicates intraquarter. IQ change-weights are based on the sum of all absolute overnight holding changes during the quarter (only for CRSP matched holdings) and thus reflect a daily rebalancing investment strategy. *Interpretation:* RH traders gravitated towards a portfolio of shares with quoted bid-ask spreads somehwere between the value-weighted and the equal-weighted portfolio.

	Median		Mean	
	Q1 2020	Q2 2020	Q1 2020	Q2 2020
Half Quoted Bid-Ask Spread, in bp/$				
RH Holding Weighted	5.6	5.8	16.3	15.5
RH IQ-Change Weighted	4.1	4.0	13.2	10.8
Marketcap-Weighted	1.1	1.4	2.7	2.8
All Unweighted	8.3	9.2	39.6	39.1
Half Quoted Bid-Ask Spread, Rate/100s				
RH Holding Weighted	$0.50	$0.50	$2.16	$2.71
RH IQ-Change Weighted	$0.50	$0.50	$2.08	$2.15
Marketcap-Weighted	$0.50	$1.00	$4.65	$5.88
All Unweighted	$1.00	$1.00	$5.09	$4.99
Stock Price, in $				
RH Holding Weighted	19.11	17.48	104.90	225.81
RH Q-Change Weighted	12.95	12.22	106.20	471.06
Marketcap-weighted	114.83	114.65	2,646.42	2,284.37
All Unweighted	17.58	15.05	128.33	120.80

Appendix: Additional Results

Table A.I

Equivalent Correlations in the Barber and Odean (2000) 1991 to 1995 Data

The table shows portfolio weight and return correlations in the Barber and Odean (2000) data for 304,929 unique end-of-month holdings from 1991 to 1996 (equities only). The "*ABO*" portfolio is based on the number of investors in each stock (on a given month-end) and equivalent to the *ARH* portfolio. *QBO* is the rolling 12-month volume and dollar-volume-based portfolio, equivalent to *QRH*. The "*ABO*xP" is similar to the *ABO* portfolio, except it first multiplies the number of investors in each stock by the stock price. The HH-VW portfolio simply adds up the dollar investment amounts over all investors. The HH-EW first normalizes each household to $1 before adding up. The most interesting correlations are indicated by a †. Panel A is for 20,949 year-end investment weight correlations. In Panel B, lines 3 and 4 are the the residual returns from a 59-month Fama and French (2015) five-factor model. *Interpretation:* Panel A: The *ABO* portfolio is similar to an equal-weighted portfolio of households. Panel B: The predictive properties of the *QBO* portfolio for the *ABO* portfolio are similar to the predictive properties of the *QRH* portfolio for the *ARH* portfolio.

Panel A: Investment Weights Correlations

		Households	
	QBO	HH-VW	HH-EW
ABO	0.75	0.78	0.97†
ABOxP	0.43	0.34	0.57
QBO		0.60	0.72†

Panel B. Portfolio Return Correlations over 59 Months

		With BO Info		BO Household		Placebo		
		QBO	ABO×P	HH-VW	HH-EW	XMKT	SMB	HML
Raw Return	1. ABO	0.963†	0.854	0.926	0.993†	0.832	0.474	–0.175
(TS correlation)	2. QBO	1.000	0.909	0.827	0.979	0.895	0.433	–0.324
5-F Residuals	3. ABO	0.815†	0.522	0.962	0.982†	0	0	0
(TS correlation)	4. QBO	1.000	0.519	0.784	0.843	0	0	0

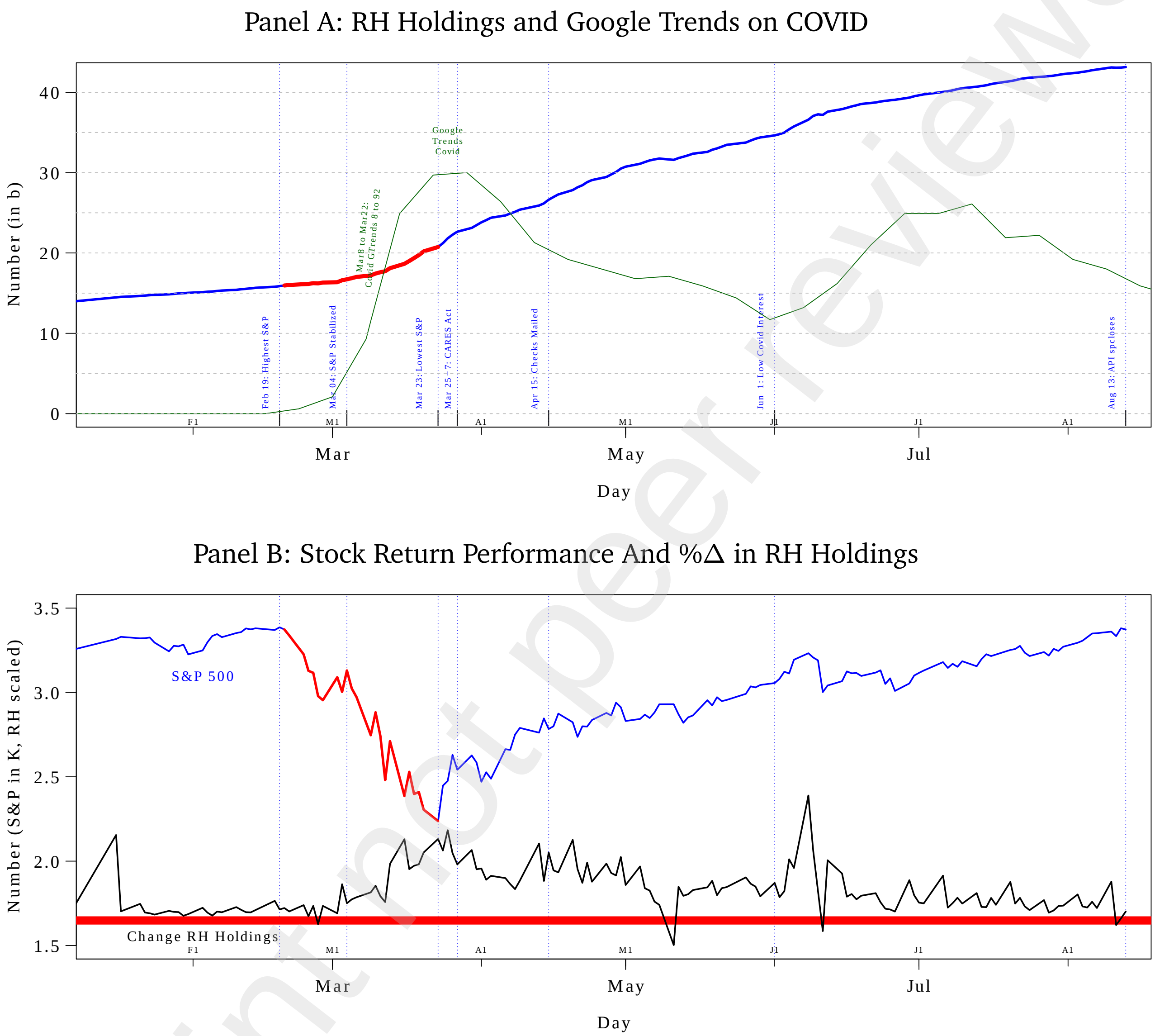

Figure A1. Sum-Total RH holdings and stock market performance in 2020. This figure is a closeup of Figure 2. *Interpretation:* The COVID-related drop in the stock market coincided with an acceleration of sum-total RH positions.