NBER WORKING PAPER SERIES

RETAIL RAW: WISDOM OF THE ROBINHOOD CROWD AND THE COVID CRISIS

Ivo Welch

Working Paper 27866
http://www.nber.org/papers/w27866

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2020

Retail Raw: Wisdom of the Robinhood Crowd and the Covid Crisis
Ivo Welch
NBER Working Paper No. 27866
September 2020
JEL No. D9,G11,G4

ABSTRACT

Small retail investors at the Robinhood (RH) retail brokerage firm from 2018 to 2020 shared with Finnish and larger US investors from the 1990s a preference for extreme recent winners and losers. Interestingly, this preference held even for the overall stock market during the March-2020 Covid crisis, indicating an absence of panic and margin calls. Thus, RH investors acted as a (small) market-stabilizing force. They were also unusually interested in some "experience" stocks (e.g., Cannabis stocks). Nevertheless, the narrative of pure irrational exuberance is misleading. Collectively, RH investors bought and held stocks with large past share-volume and dollar-volume, making them invest overwhelmingly in large rather than in obscure stocks. A portfolio constructed on the basis of just these two variables can make it possible to mimick the investments of RH traders, plausibly even beyond the sample. The collective RH crowd portfolio also did not underperform with respect to standard academic benchmark models.

Ivo Welch
Anderson School at UCLA (C519)
110 Westwood Place (951481)
Los Angeles, CA 90095-1482
and NBER
ivo.welch@anderson.ucla.edu

The online retail brokerage company Robinhood (RH) was founded in 2018 based on a plan to make it easier and cheaper for small investors to participate in the stock and option markets. RH has never charged brokerage fees, making it possible to buy and sell single shares of stocks.[1] Instead, RH earns its revenues through margin fees and cash balance interest, payment-for-order flow, and sales of its investor data to more sophisticated high-frequency traders. RH's also appealed with many other small technological innovations, such as mobile-first user interface.

As of mid-2020, RH had attracted a clientele of over 13 million investors—widely believed to be small, young, computer-savvy but also mostly first-time novice investors. "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" (WSJ, Sep 12, 2020). The website brokerage-review.com estimated that the average account size at RH was only $2,000. By August of 2020, RH had raised another $200 million of fresh capital, boosting its valuation to $11.2 billion. It is widely considered a disruptive force in US investing.

RH also offered an API from mid-2018 to mid-2020 that made it possible to obtain the (anonymous) number of RH investors holding a particular stock at that particular moment. In turn, the website robintrack.net wrote some scripts to continuously pull down the RH information (at a speed of about 20 stocks per second) and then reposted the data online with RH's blessing.

My paper investigates the history of the numbers of RH investor holdings. It first documents that the small RH investors of the 2018-20s shared some of the behavioral traits of retail investors first observed in other contexts. In particular, Grinblatt and Keloharju (2001) and Barber and Odean (2008) had shown that other retail investors in the 1990s bought stocks that had recently gone up or gone down a lot. This could be due to stocks

[1]Presumably, with their payment for order flow, RH spread the eliminated fixed-cost component into higher per-share variable costs—making it cheaper to buy and sell small positions and more expensive to buy and sell larger positions. However, given the lack of data on execution costs by *all* retail brokers in the market, and a criminal law related to price manipulation against testing execution quality with sample roundtrip trades, the relative execution quality of RH is difficult to ascertain. RH's rapid growth led other brokers to abandon brokerage fees by October 2019. It was also reputed to have played a role in inducing the merger between Charles Schwab and Ameritrade.

catching the attention of investors[2] or investors deliberately seeking of sensation. Barber and Odean (2013) surveys the behavioral literature on individual investors.

In addition to establishing that this was also the case here for individual stocks that underperformed or outperformed the market dramatically on a specific day, my paper can show that this behavior extended to the sharp market-wide Covid decline in March 2020. RH investors did not panic or experience margin calls. Instead, there is evidence that as the stock market declined, investors actively added cash to fund purchases of more stocks. Their first purchasing spike occured as early as the next day, presumably reflecting account purchasing power. The second spike occurred 3-4 days after a large market movement, roughly the time that it takes to complete a cash bank transfer. Thus, the evidence suggests that RH investors collectively acted as a (small) market-stabilizing force.

There is plenty of opportunity to poke fun at their holdings. For example, RH investors overweighted certain stocks that seemed to appeal to their interests—Ford (but not GM), Facebook in 1998 (but not in 2000), airline stocks in 2000 (but not in 1998). AMD, Snapchat, and other Cannabis stocks were also unusually popular among RH investors.

Nevertheless, although RH-type investors may very well have played a role in the demand for many otherwise obscure cannabis stocks, the RH actual crowd portfolio (ARH) was not as crazy as these "anecdotal holdings" would suggest. Instead, most of the holding interest of RH investors revolved around larger and highly liquid firms. Two readily available stock attributes—share trading volume and dollar trading volume over the last year—can explain about 50-60% of the ARH crowd portfolio's investment weights.[3] I can speculate that the remaining 40% relate to the visibility of products and stocks for my target investor group. Unfortunately, there are no readily available long time-series that would make it easy to measure this. The correlation of investment weights (of this two-attribute quasi-RH portfolio [QRH] with the actual crowd investment portfolio [ARH]) is also enough to explain about 2/3 of the rate-of-return residuals from the 5-F model in Fama and French (2015). This is similar to if not higher than the association of 5-F residuals between the

[2] See also Ben-Rephael, Da, and Israelsen (2017), DellaVigna and Pollet (2009), Fang and Peress (2009), Fang and Peress (2009), Hirshleifer, Lim, and Teoh (2009), Peng and Xiong (2006), Da, Engelberg, and Gao (2011), DellaVigna and Pollet (2009).

[3] This understates the sanity of the RH portfolio. RH investors also held better-diversified ETFs, but these were not part of the study here.

equal-weighted, value-weighted, and S&P market portfolios. The QRH is not a perfect but a reasonable proxy of the ARH portfolio.

Yet it is perhaps most surprising that the ARH portfolio, based on the collective wisdom of the crowd, did not underperform. This was the case for the 0-factor model (i.e., returns above the risk-free rate), the 1-factor model (i.e., abnormal returns adjusted for market-beta), and the 5-factor model. The alphas were positive—-and, despite the very short sample period, even statistically significant for the 5-factor model with a respectable abnormal rate of return of 1.3% per month. The QRH portfolio did not underperform over a much longer horizon, starting in 1980, either.

I Background

Wikipedia describes Robinhood Markets, Inc., as an "American financial services company headquartered in Menlo Park, California. The company offers a mobile app and website that offer people the ability to invest in stocks, ETFs, and options through Robinhood Financial and crypto trading through Robinhood Crypto. Robinhood operates a website and mobile apps for iPhone, Apple Watch, and Android. The company has no storefront branches and operates entirely online without fees. Robinhood is a FINRA regulated broker-dealer, registered with the U.S. Securities and Exchange Commission, and is a member of the Securities Investor Protection Corporation. The company's main source of revenue comes from interest earned on customers' cash balances, selling order information to high-frequency traders (a practice for which SEC opened a probe into the company in September 2020) and margin lending. The company has 13 million users." RH also had its fair share of controversies, relating to service outages, cryptocurrency, banking, payment for order flow, a security breach, and even the suicide of one of its investors.[4]

Robintrack.net (RT) was created in 2018. For about three years, RT regularly and irregularly ran a script to download all RH data made publicly available on RH's API. RH terminated its public API in August 2018 and RT froze its operations. By this time, the data base had accumulated into 3.5GB worth of data. After removing repeated

[4] Fong, Gallagher, and Lee (2014) found that discount brokerage firms had less informative trades than full-service brokerage firms. My paper tests whether their base finding (low informativeness) also holds for Robinhood investors.

intra-hour observations and unchanging holdings, it contained about 12 million ticker-hour observations. For each stock, I extracted the last UTC observation for each day. (Conveniently, the NYSE closes at 4:30pm, which is 11:30pm UTC.)

This resulted in 5,777,002 RH ticker-day observations from 802 unique days and 8,597 useful tickers.[5] Of the 8,560 RH tickers, 8,387 tickers were matchable to CRSP. Of these, 3,834 had sharecode 10 or sharecode 11. My paper focuses only on this set.

Some tickers do not appear at all or appear late in the RT data. Early versions of the script probably omitted dual-class tickers ending with '.A' and '.B', most prominently Berkshire Hathway. RT remedied this with an upgrade on 2020/01/16.[6] Nevertheless, some stocks are not in the RT dataset at all for reasons unknown, most prominently CELG (Celgene) and TWX (Time-Warner).

From 2018/05/02 (the incept day on RT) to 2020/06/30 (the end day of my CRSP data set), there were 545 valid CRSP trading days. RH suffered some systemwide outages on 2020/03/02 and 2020/03/09. The RT script should have been but was not run on 2018/08/09, on 2019/01/24-29 (4 days), and 2020/01/07-15 (7 days). The last followed the dual-class script update. This left 533 valid trading days, which are the basis for my analysis.

Over time, the number of RH investors increased, and with the law-of-large-numbers, presumably the reliability of the number of RH investors holding individual stocks. On 2018/06/29, the RH data covered only 2,947 stocks that could be matched to 3,635 CRSP stocks. On 2019/12/31, the respective numbers were 3,522 and 3,613.

The only other academic paper that I am aware of that has studied RT data is Moss, Naughton, and Wang (2020), which shows that RH investors did not care much about ESG, contrary to some experimental studies. A contemporaneous and independent research report by Cheng, Murphy, and Kolanovic (2020) studies purchases and sales of RH data, focusing primarily on *changes* in RH holdings following earlier Barber-Odean studies. In contrast, my own paper focuses more on the (level) investment weights of an inferred crowd portfolio—i.e., it is more level than change-focused. I will note overlap below.

[5] The data contained some non-sensical tickers, such as _OUT, _PRN, MTL-, PKD~ (and its sibling PKD), which I hand-removed.

[6] Thus, BRK.B suddenly appeared with 38,023 users (BRK.A with 134 users). (Figure 3 shows that BRK.A was right on the fitted line on 2020/06/30.) Other noteworthy dual-class examples, also appearing on 2020/01/16, included Royal-Shell Dutch and Lions Gate Entertainment.

Before offering an analysis, it is important to offer appropriate caveats. RH investors are a small and perhaps unusual part of the overall retail investor clientele. Their behavior may or may not be representative of the behavior of retail investors in general.

Without access to the investment amounts that individual RH investors held or transacted, my paper can investigate only the aggregate and relative stock-specific holding patterns. In particular, my paper can focus only on (1) changes in the number of RH holders for an individual stock over time; (2) changes in and the performance of an investment portfolio that is formed based on the relative number of holders. I define this Actual RH Portfolio ("ARH") as

$$w_{i,t}^{\mathrm{ARH}} \equiv \frac{n_{i,t}}{\sum_i n_{i,t}} \tag{1}$$

where i is a stock index and t a day index. ARH assumes that each investor holding represents an equal amount of dollars. For example, if RH posts twice as many owners of stock A compared to stock B, the weight of A in the ARH portfolio is twice that of B. The number of investors in each stock is unlikely to be a good approximation for shares in a value-weighted portfolio of RH or retail investors in general. It probably overweights small investors.

Importantly, this ARH portfolio is not the only one that could be considered. One of its drawback is that when a stock increases in value, its weight does not increase and the ARH portfolio does not tilt more towards it. An alternative crowd portfolio would be a portfolio in which each investor (holding) represents an equal number of shares. In this case, $w_{i,t} \equiv (n_{i,t} \cdot P_{i,t})/(\sum_i n_{i,t} \cdot P_{i,t})$. A drawback of this alternative portfolio would be that any variable that is correlated with price (such as marketcap or dollar trading volume or simply price itself) becomes nearly mechanically correlated with this portfolio's investment weights. That is, the portfolio investment weight would no longer be primarily a measure of information from RH.

The performance of this ARH portfolio is not likely to be representative of the performance of the average RH investor. In particular, few if any RH investors are likely to hold the widely-spread ARH portfolio. Instead, individual RH investors are more likely to suffer (or enjoy) diversifiable risk. NOTE THAT THIS IS DIFFERENT FROM Barber and Odean (2013).

This particular ARH crowd portfolio considered invests more when the number of investors holding a stock is higher.[7]

Multiple stock holdings could also increase suddenly when investors transfer brokerage accounts from elsewhere to RH. However, transferring assets from another brokerage firm to RH does not seem to be common. My repeated email interactions suggested that my customer-service representative did not even understand my question of how to do this. Instead, she repeatedly suggested funding the account by linking to a bank cash account.[8] It is also not possible to transfer fixed amounts of funds (e.g., via paypal or check) to fund an account. In any case, my paper focuses more on questions about ARH investment levels rather than questions about simultaneous ARH changes. Thus, it does not even matter whether investors had purchased stocks earlier and just transferred them or whether they purchased them anew.

Another source of noise is that RH gives each investor one free randomly-chosen share upon signup or referral. 2% of these new investors are "winners," in that they receive one share that is drawn from one of six stocks explicitly named on RH's bonus page with stock prices above $10. The shares given to the other 98% are impossible to ascertain. 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 this should not matter to RH investors, it can influence the reported holdings. I suspect that Robinhood merely reassigns some random share just sold by another investor to the new investor, thereby saving all external costs. If this is correct, then the sale of a larger position could easily splinter into more holders rather than fewer. It is also likely that many investors simply hold on to their share, despite the near-zero marginal trading-out cost. Thus, this holding could persist over time and effect even portfolio levels. This is noise in my study, tilting it against finding any effect.

[7] A similar assumption would be required for changes in holdings. If investors diffuse their portfolios from concentrated positions, more RH investors will be holding each stock without a necessary overall increase in investment dollars. Nevertheless, though possible, it seems implausible that the amount invested by RH in a particular stock would not increase when the number of investors in this stock increases.

[8] To fund their accounts, 13 million RH investors were willing to hand their user credentials including passwords for their bank accounts to an intermediary named "PLAID," that links Robinhood to these banks. If PLAID were to be hacked, the consequences for these investors could be disastrous. All banks advise their customers not to share their account credentials and will contest any liability if bank account funds disappear through such a channel.

In sum, the ARH portfolio considered in my paper is one feasible portfolio that was investable in-time. Its investment weights summarizes the RH crowd participation in one particular fashion.

II Some Behavioral Patterns in Changes

This section first examines *changes* in RH holdings as background evidence. It investigates whether some of the behavior patterns of retail investors documented in Grinblatt and Keloharju (2001) and in Barber and Odean (2008) also apply in my sample. The Grinblatt and Keloharju (2001) data was for Finish investors from 1994 to 1997. The Barber and Odean (2008) data was stitched together from various clienteles for U.S. investors from 1991 to 1999.

Causality is local. It is not impossible for my investors to behave differently. Neither the time period nor the types of investors overlap. My sample consists of smaller investors 25 years later, perhaps even wiser after the 2000 Tech collapse. It also includes the 2020 Covid experience—the most severe contraction of the economy since the Great Depression, all in the midst of a sharp stock market decline followed by a sharp bull stock market. Thus, it requires empirical evidence to find out whether RH investor behavior was similar or different.

A Active Holding Changes

[Insert Table 1 here: **Extreme One-Day Increases in RH Holdings**]

To help gain an intuitive understanding of the data and behavior, Tables 1 and 2 describe the most extreme cases of changes in the ARH investment weight, i.e., in a portfolio formed in accordance with the number of RH investors holding each stock.

Table 1 lists stocks with unusual one-day increases in investor interest. The common feature seems to be highly unusual one-day stock price increases or decreases.[9]

[9]Note that a stock split does not mechanically change the number of investors in a stock.

[Insert Table 2 here: **Extreme One-Day Decreases in RH Holdings**]

Table 2 lists stocks with unusual one-day decreases. There are again some rather stark stock price changes, tilting more towards losses. Sounding somewhat trite, IGC investors seem to have decided to realize their profits after a stark price increase a few days earlier. However, I could not detect rhyme or reason in these declining sentiment changes.

B Individual Contrarianism and Sensation-Seeking

[Insert Figure 1 here: **RH Holding Changes by Net-of-Market Daily Return**]

Figure 1 investigates whether RH investors were "sensation-seeking" in the Barber and Odean (2008) sense. I first categorized all trading days by their rate of return net of the market (both from CRSP). The plots then tabulate *contemporaneous* RH change statistics. No attempt is made to distinguish between causal and correlative-only associations. However, it would seem unlikely that small RH investors alone would be influential enough to cause major price changes. Similarly, no attempt is made to distinguish between anticipatory and consequent purchases. It seems even less plausible that small retail investors would collectively have had advance knowledge.[10]

The top plot tabulates the fraction of stocks that experienced increases in their RH holdings vs. decreases in their RH holdings on each trading-day. It shows that when stocks either increased or decreased by about 30% on one day, the number of RH holdings increased in about 4 out of 5 times. This contrasts with stock-days that had non-descript performances, in which the number of increases and decreases was about even. The plot also shows (in lighter colors) lines based on pre-2020 and post-2020 data. The observed pattern was stable over time.[11]

The bottom figure tabulates ARH portfolio weight changes. This takes into account both that an increase from 50 to 51 holders is not the same as an increase from 1 to 100 holders, and that other stocks on the same days may also have experienced changes. The response is still a U-shape, but the plot shows that the increase in RH investors was stronger

[10] These questions are investigated in a different data set in Barber, Odean, and Zhu (2008).

[11] Kaniel, Saar, and Titman (2007) use an order-based proxy for individual investor trading. The pattern then changes somewhat. Investors are still contrarian over monthly horizons, but they did not purchase (extreme) winners. Surprisingly, retail investors' returns were quite good over the subsequent month.

on the positive than on the negative side. This suggests that increases of interest on the down-side were often in smaller stocks, where just a few net purchasers would increase the total number of RH investors.

Cheng, Murphy, and Kolanovic (2020) similarly examine the effect of price changes and find the same RH response patterns. In sum, the evidence suggests that the "sensation-seeking" evidence in Grinblatt and Keloharju (2001)) and in Barber and Odean (2008) also appears in RH investors. It is proportionally weaker on large stock price decreases than it is on large stock price increases.

C Full Sample, Systematic Contrarianism

Retail investor could behave differently in stark and/or prolonged market-wide downturns. They could start to panic or they could face margin calls. Neither Grinblatt and Keloharju (2001) nor Barber and Odean (2008) examined this question. There simply was no large decline in the Grinblatt sample; and the biggest decline in the Barber sample was only 15% (July 1998). In contrast, my sample contains an interesting episode. After multiple days of decline in March 2020, the stock market had dropped by about 1/3, with reasonable expectations of a calamitous economic depression caused by Covid looming ahead. As of September 2020, although the economic depression has materialized, the stock market has recovered and reached new all-time highs.

[Insert Figure 2 here: **Sum-Total RH Holdings and Stock Market Performance, 2018-2020**]

Interpreting the aggregate time-series behavior is complicated by the fact that the Robinhood brokerage firm was not stable but experienced good growth throughout the entire sample period. Figure 2 shows a full-sample tally of the number of holdings. The sum-total of investor holdings grew steadily from 2018 to 2019—and then dramatically accelerated in 2020.

[Insert Figure 3 here: **Sum-Total RH Holdings and Stock Market Performance in 2020**]

The plot in Figure 3 is a closeup view. It shows that the most dramatic acceleration in holdings by RH investors appeared roughly contemporaneously with the stark decline in the stock market.[12]

[Insert Figure 4 here: **2018-2020 RH Systematic Contrarianism By Horizon**]

Figure 4 presents the same information but for different horizons in an x-y graph, plotting the performance of the S&P 500 and the percent change in RH sum-total holdings. It shows that over horizons from 1 day to 1 month, both large decreases and increases in the S&P 500 associated with large increases in the sum-total number of holdings by RH investors.

[Insert Table 3 here: **Daily Innovations in the Sum-Total Number of RH Holdings**]

Table 3 investigates the association between daily changes in the S&P 500 and daily innovations 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).

The estimated coefficients suggest a first spike in innovations in RH holdings exactly one day after the stock market increased or decreased. This suggests that RH investors learn about strong market movements towards the end of the day and purchased further holdings the following day with the remaining (margin) purchasing balance. The 0.13 up-coefficient suggests that a 10% market increase resulted in a 1.3% increase in the number of RH holdings; the 0.04 down-coefficient suggests that a 10% market decrease resulted in an 0.4% increase.[13] A second spike occurred about 3-4 days later—roughly the time that it takes for a bank transfer to complete. This second spike is more symmetric, with strong bull and bear markets having an equal influence on innovations in RH holdings. The two-day coefficient of about 0.1 suggests another 1% increase—this one almost surely funded by dollars previously held in cash or savings accounts and now flowing into the riskier equity markets. Not shown in the regressions, there is also some evidence that lower volatility and returns in bitcoin increased RH participation. However, this effect is smaller.

[12] A spike in January hints at tax-related, bonus-related, or New-Year-resolution related purchases.

[13] The stock return is negated in this variable to make the interpretation easier.

In sum, just like strong price increases or decreases of individual stocks can explain additional individual RH purchases of a stock, strongly positive or strongly negative overall market movements can predict/explain unexpected increases in RH positions. This evidence suggests that RH investors did not retreat from the stock market during its starkest decline. It appears that RH investors served a small but active market-stabilizing role during the Covid crash. In this case (as in the other stock-market crashes in our lifetimes), holding on and/or increasing the position would have paid off well in the subsequent stock-market boom.

D Some Odd Holdings in the RH Portfolio

[Insert Table 4 here: **Highest ARH vs VW Log-Investment-Rank Difference**]

Reading popular descriptions, the behavior of RH investors is subject to common ridicule. Indeed, RH investors ended up with some rather unusual portfolio positions. In the most extreme example where the value-weighted market cap rank was much lower than that of RH investors, Table 4 shows that RH investors were particular fans of India-Cannabis (IGC), which ranked 27th on their holdings in 2019 with 0.59% of their 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 think of a rational portfolio in which IGC would deserve an investment holding similar to that of J.P.Morgan.

Not shown, other Cannabis stocks similary attracted unusual interest by RH investors. (The RH mobile interface even has a special button dedicated to pulling up the most popular Cannabis stocks.) At the end of January 2019, Aurora Cannabis (ACB) was briefly the most widely held stock by RH investors, with 244,532 investors (AAPL being second with "only" 237,521 investors)!

Moreover, RH investors also had a relative preference for certain oil&gas-related exploration companies and bio-pharmaceutical-related companies. All these investments tended to be highly risky.[14]

In sum, it is easy to tell a tale that RH retail investors fit the stereotype of the unsophisticated gambler, earning low returns and being taken advantage of by more sophisticated professional traders. However, the next section will show that this narrative seems incorrect.

III The Overall Portfolio Perspective

Although RH investors were presumably large investors in some rather odd stocks (as in IGC), this does not mean that the most important holdings in the ARH portfolio were in these odd stocks, too. This section provides evidence that the ARH crowd portfolio was a lot more ordinary. The odd holdings described in Table 4 are not the rule but the exception. From the perspective of the ARH portfolio, they can be described as "interesting curiosity holdings."[15] As far as I am aware, studying the level holdings of retail investors has not received a lot of attention to date.

[14] Remarkably, the historical standard deviation of stock returns is not a great predictor of ARH holdings, because it is largely subsumed by trading volume in Table 7.

[15] As already mentioned in footnote 3, my analysis understates the "sanity" of ARH, because it excludes many better-diversified ETF portfolios with share codes other than 10 or 11, also commonly held by RH investors.

A Large Holdings

Figure 5 provides a graphical perspective of ARH portfolio holdings for two snapshots, one at the end of 2018, the other towards the end of the sample in mid-2020.

[Insert Figure 5 here: **Snapshot ARH Weights by Log-Rank of Market Cap**]

Although IGC was perhaps the strangest popular holding in Table 4, it was not particularly noteworthy in the overall portfolio. After all, even at its height, IGC still represented only 0.5% of the portfolio. This leaves 99.5% for other holdings. Many other stalwarts represented more than twice of IGC's weight. The outliers other than IGC from Table 4 would be even less notable.

This does not mean that the ARH portfolio mimicked the marketcap-value weighted portfolio. For example, the top plot (2018) shows that GE and F were as popular as AAPL and both exceeded the holdings in MSFT. Other popular stocks were AMD, Fitbit, GoPro, Netflix, Snapchat—firms with products familiar to computer-savvy Millenials.[16] The blue line is a near perfect fit of the value-weighted market-cap weight, by design, for comparison.

The bottom plot shows that the (blue) market-cap weighted curve had steepened by 2020, with AAPL valued at $2 trillion representing about 4% of the overall stock market. In contrast, the black RH curve for ARH weight by marketcap remained roughly unchanged. Yes, the ARH portfolio continued to invest strongly in AAPL (with about 2% weight), but less aggressively so than the value-weighted market. Instead, RH investors overweighted Disney, GE, Ford, and airlines stocks. A fair characterization is that although retail investors had loved "new-economy" stocks in 2018, their portfolio had tilted more towards fallen "old-economy" stocks by mid-2020.

[Insert Table 5 here: **ARH Investment Weights *Far Above* VW Investment Weights**]

Table 5 lists the largest holdings by RH-portfolio investment weight. With data for both snapshots for a given stock, it also provides a (possibly ex-post) perspective on their evolution. A stock can see increases both in its ranking relative to other stocks and in its marketcap-relative weight in the RH portfolio if

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

1. it is *more* actively purchased by (new) investors *relative* to other stocks;
2. its price (and thus its weight in the value-weighted market portfolio) decreases relative to other stocks.

The table shows some instances of both. For example, although American Airlines (AAL) dropped from a high of about $50/share in late 2018 to about $10/share in mid2020, reducing its weight in the value-weighted CRSP index from about 0.06% to 0.02%, RH investors adored it. The 7,300 original holders in 2018 had turned into 654,611 holders by 2020. American Airline's weight in the ARH portfolio increased from 0.12% to 2.39%. Facebook (FB) had good stock market performance, raising its weight in the value-weighted market from 1.34 to 1.78. But it fell out of relative favor with RH investors, with holdings increasing only 40% rather than the 360% that the sum-total number of RH holdings increased. PLUG (a hydrogen fuel researcher) benefitted both from strong rate of return performance and increased interest.

In sum, even though some of the investments in odd small-cap experience stocks were eye-popping, these positions represented a small part of the RH portfolio. They were a storm in a teacup. The big picture happened in larger stocks, typically more familiar to and focused on retail consumers and especially Millenials.

B The Performance of The ARH Crowd Portfolio

[Insert Table 6 here: **Return Performance of the ARH Crowd Portfolio, 2018 – mid-2020**]

Table 6 shows the next-day performance of daily rates of return for the ARH portfolio. Again, this portfolio is not representative of those of individual investors. It is more appropriately viewed as a "crowd wisdom" portfolio.

To analyze the performance of the ARH portfolio, I rely on the familiar methods of Fama and French (2015). The left side of Table 6 shows that the average daily performance of the ARH portfolio was a positive 10 basis points per day on the 0-Factor model, i.e., net of the prevailing risk-free rate (from Ibbotson via Ken French's website). This abnormal performance declines to about 5 basis points on the 1-Factor model. This is performance above the stock market rate of return when the (ex-post realized) market exposure of 1.1

is taken into account. The abnormal performance increases back to 9 basis points on the 5-Factor model. This is because the ARH portfolio had positive exposure on the SMB factor and negative exposure on the CMA factor.

The right side of Table 6 shows that when portfolios are formed at the end of each month and held until the end of the following month, the performance remains about the same. The factor exposures become less precise, perhaps explaining the change in the 5-factor model CMA exposure. The abnormal rate of return on the 5-F model is a very respectable 1.30% per month and statistically significant.

The performance regressions assume equal investments in different periods. They ignore the effects of time-varying entry and exit into the stock market. They thus understate that investors had deployed more capital just after the March 2020 decline, which would have benefitted further from correctly timing the market.

In sum, the ARH portfolio performed surprisingly well. Collectively, the RH crowd may have anticipated subsequent interest of other slower retail and other investors.[17] It also suggests that the narrative that RH investors were cannon fodder for more sophisticated day traders seems misplaced.

IV Mimicking The ARH Crowd Portfolio

A The Attributes Determining ARH Investment Weights

What stock attributes explain the investment interests of the RH clientele? The answer can allow us to explore when a strategy mimicking the behavior of RH investors would have performed well and when it would have performed poorly, even beyond the short available RH 2018-2020 sample period.

To answer this question, a few issues must be kept in mind. First, it is more important for a prediction of investment weights to explain the top 50 holdings than the bottom 3,000 holdings—and these 50 top holdings are by-and-large not odd experience micro-stocks like IGC. Mispredicting the weights of cannabis stock is forgivable. Second, the ARH portfolio

[17] Good performance by stocks purchased by retail investors over a short horizon was also the case in Barber, Odean, and Zhu (2008).

has a lot of hysteresis. The figures have shown that many weights are 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 not useful to attempt to explain changes in ARG weights, because we have neither any earlier ARH portfolio holdings nor occasional ARH holdings on which we could recalibrate predicted ARH portfolios weights. Fourth, regressions predicting ARH weights should not so much be judged by their ability to predict these weights, but by their abilities to produce portfolio returns similar to those displayed by the ARH portfolio. Thus it is not of great concern that the weight-predicting regressions below include many days for the same stock with similar weights and characteristics, or that its T-statistics would be misleading, or that the regressions are misspecified on a number of other dimensions (such as heteroskedasticity or codeterminants).[18]

[Insert Table 7 here: **Explaining ARH Investment Weights (in %)**]

With 1.7 million ticker-day observations, it is feasible to start with little theory and try to disentangle similar variables empirically. Table 7 begins with a large panel regressions with many variables in the pre-2020 sample, allowing the post-2020 sample to be a hold-out sample. This "kitchen sink" regression contains a set of variables inspired by earlier (largely behavioral) research, as well as some more ad-hoc variables:

IPO Variables: IPO<1year, IPO<3years, Days Since IPO logged, days since IPO;

Name: A dummy when the ticker starts with the letter "A";

Share Volume, Dollar Volume, Marketcap: Basic value, a "transformed weight" (see explanation below), a rank of weight, log of weight, squared weight; 5-day, 22-day, 252-day, and 504 day average weights, calculated as mean own value divided by mean of total-across-stocks value; historical mean dollar trading volume divided by current price (thus adjusting for stock splits);

Further Marketcap: Dummies for being among the top-500 or top-100 stocks by marketcap; a nonlinear option-like transform (500 - rank for big firms, 0 beyond);

[18] The measures in Lee and Ready (1991), Kelley and Tetlock (2013), Boehmer et al. (2019) track retail investor order *flow*. This is likely to be related to *changes* in investment weights, not to *levels*. It is simply a different and mostly unrelated measure.

Share-Price: Plain, squared, and log; stock split measures; Price is below \$1, \$10, \$50, or \$100.

Min-Max Distance: the percent Distance to All-time Max, to 252-Day Max, to All-time Min, to 252-Day Min; all plain and squared;

Various: the largest-mcap stock in the same month, the number of stocks in month, the crsp day index, and even the permno as a placebo.

The "transformed weight" input variables will play an important role and require explanation. They map a raw variable into a variable that mimicks a portfolio weight, such that the cross-sectional sum adds to 1.0 on each day. The calculation is equivalent to that used for value-weighted portfolios. For example, if stocks A-D have attributes 2, 4, 6, and 8, the investment weights would be 10%, 20%, 30% and 40%. The advantage of such weight-normalized variables is that their units are easily relatable to the dependent variable, the ARH portfolio investment weight.

The fit can also be improved with a later-stage equivalent portfolio-weight procedure. The OLS regression model fits some negative weights and the fitted portfolio weights would no longer add to 1. Thus, to assess the quality of the regressions, I truncate fitted values to 0 and then rescaling the weights. The reported R^2 (and portfolio returns below) are based on these rescaled weights.

There are other more sophisticated econometric techniques that could be used to allow the statistical model to enforce these restrictions itself. However, this would make it more difficult to replicate the analysis and be subject to a "specification search" critique. The simplicity of the linear procedure with adjustment here is appealing.

The first row in Panel A shows that the pre-2020 kitchen-sink regression can explain about 66.60% of the variation in ARH portfolio weights. However, it is not likely that such an overfitted highly-collinear model would remain stable over other sample periods. It is useful primarily to identify important variables (two stick out) and to learn the limit of how well a model with the employed variables could fit.

A simpler model is more likely to be more stable and thus more useful out-of-sample. I therefore employed a "forward-selection" regression procedure on the kitchen-sink variables.

The second row shows that a two-variable model can explain 58.81% of the variation in ARH weights:

1. The mean share trading volume over the last year for each stock, divided by the mean total share trading volume (transformed into a portfolio weight);
2. the same for dollar trading instead of share trading volume.

The coefficient estimate on the former is about twice that of the coefficient estimate on the latter. I shall refer to the (renormalized) weights from this model as the "Quasi RobinHood" ("QRH") portfolio.

Its two input variables are mechanically correlated with one another. The dollar-volume is also nearly mechanically correlated with the not-included marketcap. Moreover, empirically, these two variables also correlate well with many other omitted variables such as return volatility and distance to min/max price.

Two alternative specifications are similar and nearly as good. A first does not use a full year's worth of trading, but a shorter period (merely a month). A second does not use the mean number of shares trading, but the mean dollar-volume of shares trading divided by the last price. (This is a sensible adjustment for stock splits.)

When explaining ARH weights with the QRH model, it is true that there are still many other variables that could add statistically significant power. The table shows the performance of a few of them (based on intuition-augmented forward selection). On the margin relative to the QRH variables, the ARH portfolio still underweights the biggest 100 stocks, the biggest 500 stocks, and younger firms. However, these are correlations merely on the margin, with most of their (unconditional) positive relation having already been captured by the two trading variables. Most importantly, their marginal increase in fitting the observed ARH weights is modest.[19]

Panel B shows that all models perform worse in the post-2020 holdout sample—perhaps not surprising given the continued large inflow of RH investors and the turbulent market conditions. The kitchen-sink model's explanatory power declines to 61.87%. The QRH model explains only 49.45%. The larger 12% difference relates to primarily to shorter-term (say, weekly) trading volume, which was not important in the 2018-9 sample but did

[19]Not shown, the rate of return of these models is also just a few basis points better.

become more important in 2020. As to the two coefficients themselves, the past-year share trading volume gains in importance relative to the past-year dollar trading volume.

Panel C shows that a model that fits to the entire sample is similar to the 2018-9 sample, if only because it shares most of its observations with the former.

[Insert Table 8 here: **Implication of QRH Model Differences For QRH Investment Weights**]

Even when models (and their coefficients) look different, it is difficult to assess whether their predictions are similar. Panel A of Table 8 shows that differences in the ultimate weight predictions between using the pre-2020 coefficients or the post-2020 coefficients are small. The fitted weight correlations from the two models are always 98% or higher, regardless of sample.

Panel B shows that this similarity does not necessarily extend to the difference between the kitchen-sink and the two-variable models. For example, on June 30, 20202, the fitted investment weights from two models had only 87% correlation in this pure cross-section. Again, the correlation was better in the 2018-9 sample than it was more recently.

In sum, the ARH portfolio tilts heavily towards stocks with a lot of trading activity over the past year. Other variables, even when statistically significant, don't add a lot of explanatory power in fitted values to the QRH model portfolio.

I can speculate that the remaining 40% unexplained variation in investment weights relates to stock attributes for which there are no readily available (long-term) comprehensive proxies. These would be variables such as excitement about or visibility of products by retail customers, measures of the share of consumer expenditures spent on these companies, their brand values and advertising, the excitement and promise of new technologies, and/or perhaps even their recent performance. Such variables could potentially help explain why RH investors liked Delta and American Airlines more than United Airlines; why they liked Ford better than General Motors; why they liked Disney more in 2020 than in 2018; or why they liked Facebook more in 2018 than in 2020. This investigation is left to future research.

B Comparing Portfolio Performance

I can now evaluate the performance of a portfolio that is inspired by—but not the same as—the ARH portfolio. This QRH portfolio tilts towards high volume and high dollar-volume stocks, in accordance with the QRH model from Table 7. As already noted, its fitted investment weights are truncated to be above zero, and its holdings are rescaled to sum to 1 first (before the QRH portfolio rate of return can be calculated).

[Insert Table 9 here: **ARH and QRH Portfolio Return Correlations, 2018 – mid-2020**]

Table 9 shows that it does not matter whether returns are calculated from coefficients based on the pre-2020 or the post-2020 sample. In particular, I calculated one full series with "IS" coefficients, where each investment weight is based on the model in the same period, and one full series with "OOS" coefficients, where each investment weight is based on the model *not* estimated in the same sample period. (That is, I use the pre-2020 model coefficients on the post-2020 stock attributes and vice-versa.) The two return series had 99.99% correlation with one another.

Panel B shows that there is some further slippage in association when the universe is expanded to include all CRSP stocks and not just the stocks in the RH sample. (Adding about 300 albeit often smaller stocks, this yields a slightly different rate of return series.) The correlation between the two QRH series is an imperfect but unconcerning 99.78%.

Panel C shows that much of the 97.82% correlation between the actual and the quasi RH portfolio rates of returns was due to shared exposure to the market-factor. After subtracting out the market rate of return, the correlation drops to a respectable but much lower 85%. Pushing this return-factor analysis one step further, the correlation between residuals from the ARH and the QRH models on the 5-F benchmark model is about 79%.

To put this 79% figure in perspective, the 5-F residuals between the CRSP S&P-500 rate of return and the value-weighted rate of return is only 27% in the same time-period. The correlation of the residuals for the equal- and value-weighted market-portfolio is only 72%. There is more association between the ARH and the QRH portfolio than there is between common but different overall stock market portfolios.

In sum, the QRH portfolio return series is a good but not a perfect substitute for the ARH portfolio return series.

[Insert Table 10 here: **Return Performance of the QRH Portfolio, 2018 – mid-2020**]

Table 10 shows the performance of the QRH portfolio in the 2018-2020 sample. In Panel A, on a daily basis, the QRH portfolio performed well, earning an alpha of about 5.1 basis points per day net of the risk-free rate, and 4.2 basis points relative to the FF 5-factor model. Relative to the CAPM market-model, QRH earned an "on-model" 0.4% appropriate rate of return.

However, the QRH portfolio delivered about 4 basis points less than the actual ARH portfolio from Table 6, statistically significant. The RH crowd simply knew better how to invest than this naive two-attribute QRH model.

Panel B shows that the same inference obtains when trading is restricted to be once-per-month and returns are compounded over the following month. The QRH portfolio had a positive adjusted rate of return of 1.035% per month above the risk-free rate, and a positive alpha of of 0.505% on the 5-factor model. On the market-model, it lost an insignificant 9.7 basis points per month. Again, the actual ARH portfolio comfortably outperformed the naive QRH portfolio.

With a reasonable mode, we can now ask how a portfolio (of a crowd of retail investors) with QRH-like preferences would have performed over a longer horizon.

[Insert Table 11 here: **Return Performance of the QRH Portfolio, 1980 – 2020**]

Table 11 shows that from 1980 to 2020, the QRH portfolio would by-and-large have achieved a similar performance as it did from 2018 to 2020. It would have had positive excess and 5-factor abnormal return performance, with modestly negative (and statistically insignificant) abnormal 1-F market abnormal return performance.

[Insert Table 6 here: **Return Performance of the QRH Portfolio After 1980**]

Figure 6 shows the year-by-year performance of the QRH portfolio, with daily returns summed over yearly intervals. 1999 and 2009 were spectacularly good years for QRH investors, with performance about 30% better than that of the market. The worst year was 2000, the year of the collapse of the "1999-9 Tech bubble period," with underperformance of about 20% below an already dismal overall market performance. This conforms to informal anecdotes from the era.

say that this is not necessary inconsistent with poor performance in Barber and Odean (2013). different sample, different portfolio. not individual, but aggregated crowd portfolio. they find trading hurtes them.

(barber-odean-2000 have 16-months average holding period.)

barber-odean-2000 is levels. show that active-investor households did worse than buy-and-hold households. similarly GK look at buy/sell ratio to see if they make good choices. can't even reconcile them without data. also, maybe they do trade out of better stocks and these are purchased by new investors. I would not know. all of these basically are changes predicting ?long?-term returns.

kelley-tetlock show positive performance on short horizon, too. all have positive findings.

I don't have much distinction between active and passive investors—I combine them and benchmark to the models.

I can't say anything about persistence of the best guys.

no evidence of disposition effect here (but again, we are just the crowd).

lack of diversification == which is why I called this the crowd pfio, and not a representative pfio.

universe is specific—omitting many others for which I have unreliable crsp data.

V Conclusion

By and large, RH investors liked stocks with high trading activity over the last year. There were exceptions. They fell in love with some obscure experience stocks, such as cannabis stocks, and some risky technology/gamble stocks. And they displayed some of the same behavioral-finance biases documented in previous literature, especially with respect to a liking of stocks experiencing extreme stock price increases or decreases.

Nevertheless, the RH portfolio as a whole was not greatly tilted toward these stocks. Many of their obscure holdings are better described as curiosity pieces. On the whole, the RH portfolio seems more tilted away from the very largest stocks and towards stocks with whose products retail customers would be familiar with. By mid-2020, RH investors had even pivoted towards fallen old-economy stocks.

Moreover, the ARH "crowd wisdom investment portfolio" performed reasonably well, earning positive alpha with respect to the risk-free rate, the market-model, and (statistically significantly) the Fama-French 5-factor model. And ARH earned these positive abnormal returns while acting as a contrarian market-stabilizing force and without panicking during the 33% Covid drop in March 2020. This steadfastness was presumably rewarded in the market recovery, although this was not measured by my performance regressions.

Further analysis shows that a simple two-variable attribute model can explain a large part of RH investor preferences (and consequent ARH portfolio return). This QRH model calculates simple transforms of share volume and dollar volume over the last year—weighing the share volume about twice as much as the dollar volume. The resulting return factor can plausibly capture much of the RH investor preferences. Above and beyond the 5-F model, it associates with the ARH model about as well as different stock market indexes associate with one another.

Thus, my paper tracked the performance of the ARH-inspired QRH portfolio over a longer timeframe. Appropriate caution to the interpretation of this portfolio track record apply. Small retail investors could have had evolving preferences over longer sample, and/or better stock-picking abilities than this naive model—which is reinforced by the higher alpha of ARH in the overlapping sample. It also seems implausible that bigger retail investors would behave just like the "micro" RH investors. With a different age profile, Larger retail investors with more to loose may have preferred more conservative holdings—although even the RH crowd portfolio was by no means as aggressive as often claimed.

Nevertheless, until there is a better model, this simple two-variable model can act as a proxy for the investment performance of small retail investors in some contexts. It will be interesting to find out if this model will continue to predict well in future years.

References

Barber, Brad M. and Terrance Odean (2000). "Trading is hazardous to your wealth: The common stock investment performance of individual investors". In: *The journal of Finance* 55.2, pp. 773–806.

— (2008). "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors". In: *The Review of Financial Studies* 21.2, pp. 785–818.

— (2013). "The Behavior of Individual Investors". In: *Handbook of the Economics of Finance*. Elsevier. Chap. 22, pp. 1535–1565. DOI: http://dx.doi.org/10.1016/B978-0-44-459406-8.00022-6.

Barber, Brad M., Terrance Odean, and Ning Zhu (2008). "Do retail trades move markets?" In: *The Review of Financial Studies* 22.1, pp. 151–186.

Ben-Rephael, Azi, Zhi Da, and Ryan D. Israelsen (2017). "It depends on where you search: Institutional investor attention and underreaction to news". In: *The Review of Financial Studies* 30.9, pp. 3009–3047.

Boehmer, Ekkehart, Charles M. Jones, Xiaoyan Zhang, and Xinran Zhang (2019). "Tracking retail investor activity". In: *Available at SSRN 2822105*.

Cheng, Peter, Thomas J. Murphy, and Marko Kolanovic (Aug. 2020). *Follow the Robinhood Money*. Tech. rep. J.P.Morgan.

Da, Zhi, Joseph Engelberg, and Pengjie Gao (2011). "In search of attention". In: *The Journal of Finance* 66.5, pp. 1461–1499.

DellaVigna, Stefano and Joshua M. Pollet (2009). "Investor inattention and Friday earnings announcements". In: *The Journal of Finance* 64.2, pp. 709–749.

Fama, Eugene F. and Kenneth R. French (2015). "A five-factor asset pricing model". In: *Journal of Financial Economics* 116.1, pp. 1–22.

Fang, Lily and Joel Peress (2009). "Media coverage and the cross-section of stock returns". In: *The Journal of Finance* 64.5, pp. 2023–2052.

Fong, Kingsley Y.L., David R. Gallagher, and Adrian D. Lee (2014). "Individual investors and broker types". In: *Journal of Financial and Quantitative Analysis*, pp. 431–451.

Grinblatt, Mark and Matti Keloharju (2001). "What makes investors trade?" In: *The Journal of Finance* 56.2, pp. 589–616.

Hirshleifer, David, Sonya Seongyeon Lim, and Siew Hong Teoh (2009). "Driven to distraction: Extraneous events and underreaction to earnings news". In: *The Journal of Finance* 64.5, pp. 2289–2325.

Kaniel, Ron, Gideon Saar, and Sheridan Titman (2007). "Individual investor trading and stock returns". In: *Journal of Finance* 62.3, pp. 1139–1168.

Kelley, Eric K. and Paul C. Tetlock (2013). "How wise are crowds? Insights from retail orders and stock returns". In: *The Journal of Finance* 68.3, pp. 1229–1265.

Kumar, Alok (2009). "Who gambles in the stock market?" In: *The Journal of Finance* 64.4, pp. 1889–1933.

Lee, Charles M.C. and Mark J. Ready (1991). "Inferring trade direction from intraday data". In: *The Journal of Finance* 46.2, pp. 733–746.

Moss, Austin, James P. Naughton, and Clare Wang (May 2020). *The Irrelevance of ESG Disclosure to Retail Investors: Evidence from Robinhood*. Tech. rep. University of Iowa, University of Virgina, and University of Colorado.

Newey, Whitney K. and Kenneth D. West (1987). "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708. DOI: doi:10.2307/1913610.

Peng, Lin and Wei Xiong (2006). "Investor attention, overconfidence and category learning". In: *Journal of Financial Economics* 80.3, pp. 563–602.

Zarnowitz, Victor and Louis A. Lambros (1987). "Consensus and Uncertainty in Economic Prediction". In: *Journal of Political Economy* 95.3, pp. 591–621. DOI: http://www.jstor.org/stable/1831979.

Exhibits

Table 1: Extreme One-Day Increases in RH Holdings

		Rel Day; Holdings in K						ARH		stock return in percent		
Date	Ticker	–2	–1	±0	+1	+2	Δ#	w	Δw	r_{-1}	r_0	r_{+1}
2018/07/26	FB	108	114	156	166	170	42,083	3.22	0.85	0.4	–18.7	–0.1
							Good 2nd-Q earnings, but below expectations, 10% price drop					
2020/01/16	INPX	0	0	26	26	25	26,376	0.27	0.27	–55.5	8.1	–7.2
							1/8: Reverse Stock Split, IOT Sensor Product wins award, Canada Patent.					
2018/10/02	OGEN	17	27	41	39	37	14,336	0.74	0.25	38.8	123.2	–51.1
							Various minor news					
2018/07/17	NFLX	101	107	118	118	117	11,000	2.49	0.23	1.3	–5.6	–1.4
							Poor earnings report, Walmart considers competitive service.					
2019/11/01	FIT	252	252	271	275	275	18,924	3.06	0.21	5.8	14.6	–1.8
							Google agrees to buy Fitbit					
2020/03/06	INO	107	111	137	155	150	25,521	1.16	0.20	25.5	45.5	–22.6
							Accelerated Covid DNA vaccine					
2020/06/03	GNUS	49	66	112	145	141	46,014	0.46	0.19	52.6	95.9	–13.2
							Raised funding for digital children network					
2018/07/27	FB	114	156	166	170	173	9,832	3.40	0.19	–18.7	–0.1	–1.6
							See above. Also acquires Redkix.					
2020/06/09	NKLA	33	77	125	130	137	48,019	0.49	0.18	102.5	9.6	–17.9
							Various progress reports. See also below.					
2020/05/18	SRNE	16	51	92	92	90	41,302	0.41	0.18	157.6	–7.0	–15.6
							Antibody has various positive Covid vaccine news.					

Explanations: These are sample cases with relatively large one-day AHC investment weight increases.

Interpretation: There are typically large daily rates of return on the day before or the day of the increase.

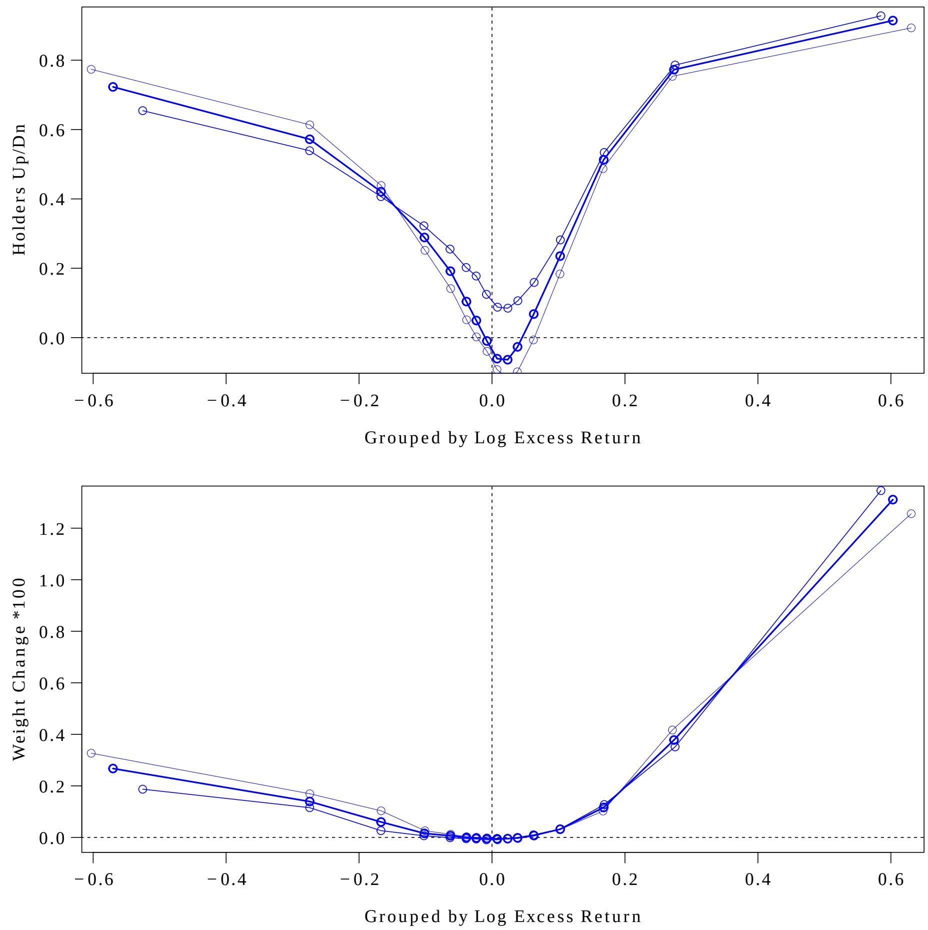
Table 2: Extreme One-Day Decreases in RH Holdings

Date	Ticker	Rel Day; Holdings in K −2	−1	±0	+1	+2	Δ#	ARH w	Δw	stock return in percent r_{-1}	r_0	r_{+1}
2019/02/27	IGC	78	78	42	42	42	–36,084	0.63	–0.71	–2.1	1.9	13.7
		Cannabis stock wins appeal to relist on NYSE. Stock rises dramatically. Q3-19 earnings release.										
2018/11/02	INPX	39	40	1	1	1	–38,583	0.03	–0.67	–3.5	–31.5	1.8
		See also above. Reverse stock split.										
2020/06/04	NKLA	95	100	21	33	77	–78,685	0.09	–0.32	6.9	–0.3	4.0
		Various progress reports. see also previous table.										
2020/02/05	TSLA	154	163	148	151	152	–15,339	1.43	–0.15	12.2	–18.3	1.6
		Good earnings report. Model-Y news. German factory. Weibo negative report.										
2020/03/13	AIKI	18	18	1	1	1	–17,240	0.01	–0.14	–8.6	–33.8	–12.0
		Name change to emphasize Alzheimer and Multiple Sklerosis drug dvlpmt										
2019/02/04	OHRP	8	8	0	0	0	–8,742	0.00	–0.14	–4.6	–23.7	–2.5
		Reverse stock split										
2018/10/01	TSLA	95	101	94	94	94	–6,876	1.69	–0.13	–13.9	17.0	–3.1
		Musk steps down as chairman and settles with SEC. Next day good earnings news.										
2018/10/04	IGC	89	83	76	72	80	–7,259	1.35	–0.13	–32.0	–26.8	–36.3
		10/2: ATM offering completed										
2020/01/16	FIT	257	257	256	257	259	–898	2.59	–0.11	0.6	–0.5	1.0
		Scripps Research claims Fitbit can detect flu										
2018/08/02	TSLA	85	85	80	77	76	–4,857	1.61	–0.11	1.0	15.7	–0.9
		OK Earnings news. Announces China plan										

Explanations: These are sample cases with relatively large one-day AHC investment weight decreases.

Interpretation: There are sometimes large daily rates of return on the day before or the day of the increase.

Figure 1: RH Holding Changes by Net-of-Market Daily Return



Explanations: Stock-days are first grouped by net-of-market stock returns into about 20 (non-equal-spaced) categories. Within each category, the y-axis presents mean statistics for the full sample (dark) and two subsamples (light, pre-2020 and post-2020). The top plot shows (a dummy for) increases vs decreases. The bottom plot shows the net change in the ARH investment weight. This weights a larger number of investor changes more (common in bigger stocks with more RH holders) and also takes other ARH changes on the same day into account.

Interpretation: RHers preferentially purchase large movers. This confirms earlier work. Contrarian increases are concentrated in small stocks with small increases relative to total number of RH holdings.

Source: 3-mkdbyret.R: 4 Sep 2020 (files d*-by-xret.pdf).

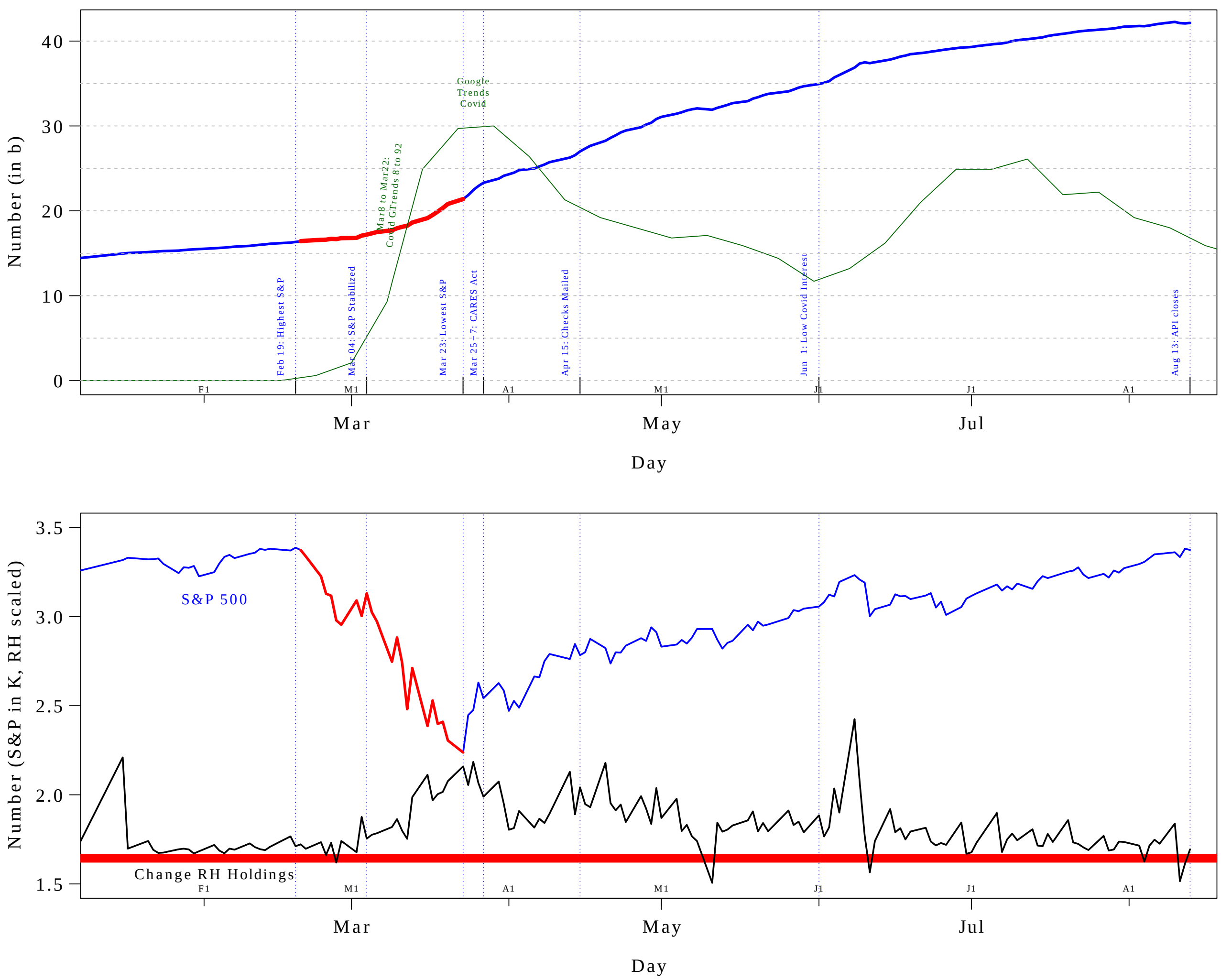
Figure 2: Sum-Total RH Holdings and Stock Market Performance, 2018-2020

Explanations: The top plot shows the total number of RH investors. The green line shows the interest in Covid on Google Trends. The bottom plot shows the S&P500 index and the percent change of the sum-total in RH holdings. The red line indicates zero growth. Figure 3 is a closeup for 2020.

Interpretation: RH investing accelerated around the beginning of the Covid crisis in the US.

Source: 4-mktsplots.R: 4 Sep 2020 (files rhs-[top||bot-2018.pdf).

Figure 3: Sum-Total RH Holdings and Stock Market Performance in 2020

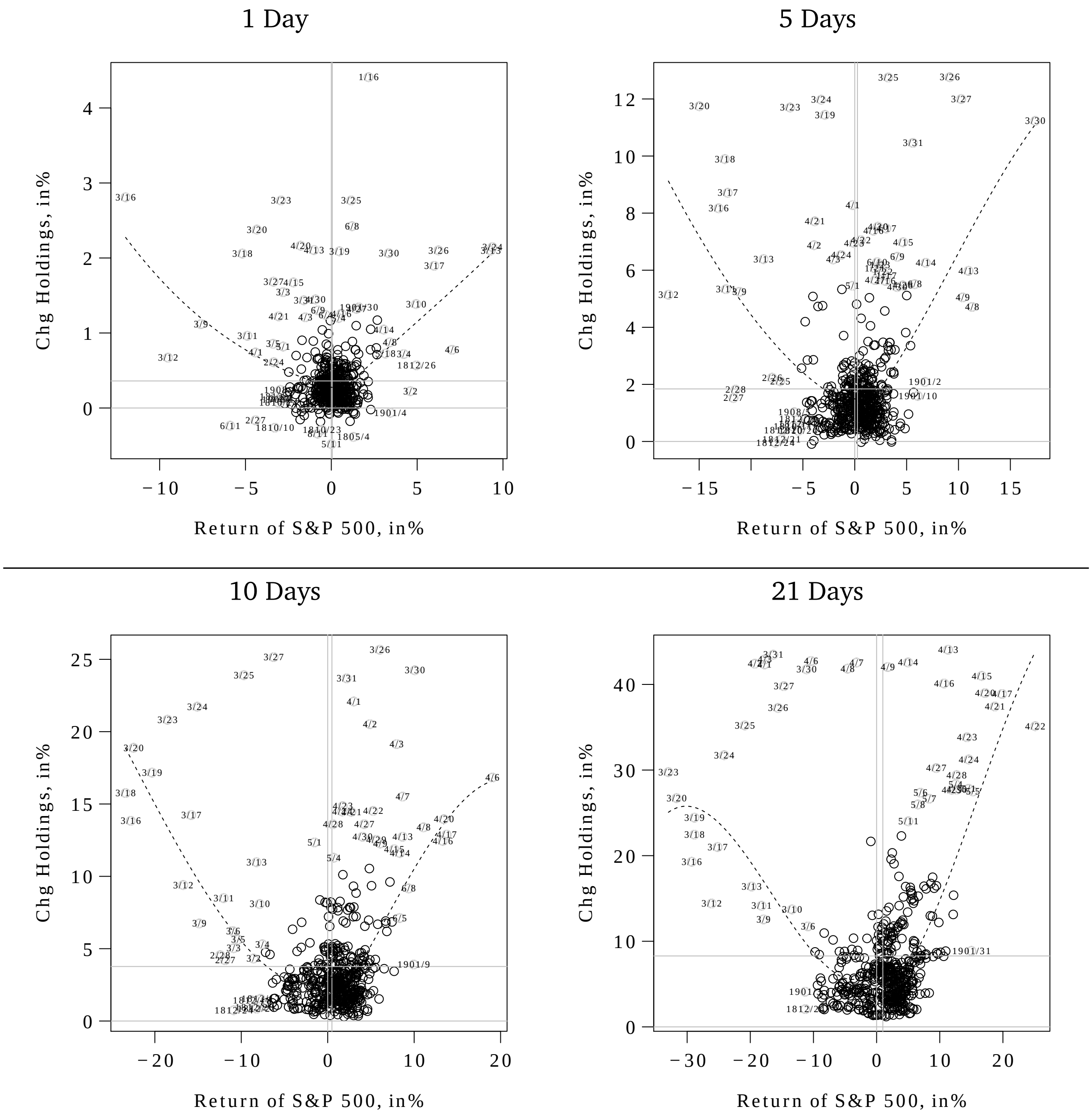


Explanations: This figure is a closeup of Figure 2.

Interpretation: The Covid-related drop in the stock market coincided with an acceleration of sum-total RH positions.

Source: 4-mktsplots.R: 4 Sep 2020 (files rhs-[top||bot-2020.pdf).

Figure 4: 2018-2020 RH Systematic Contrarianism By Horizon



Explanations: This figure shows the percent change in the sum-total of RH holdings vs. the stock-market return, similar to Table 2, but in X-Y rather than time-series format and over longer (overlapping) intervals, too. Large changes are named.

Interpretation: Large S&P500 price drops did not deter RH investors.

Source: 4-mktsplots.R: 4 Sep 2020 (files rhts-xy-*.pdf).

Table 3: Daily Innovations in the Sum-Total Number of RH Holdings

	Full Sample (D=563)				>2020/02 (D=134)			
	Coef	T	Coef	T	Coef	T	Coef	T
(Intercept)	0.08	3.20	0.03	2.74	0.06	2.04	0.03	0.92
Lagged Values								
lag(Y, 1)	0.48	5.54	0.23	3.36	0.40	6.09	0.34	5.77
lag(Y, 2)			0.26	10.08	0.17	2.57	0.16	4.05
lag(Y, 3)			0.18	4.25	0.14	2.34	–0.06	-0.86
lag(Y, 4)							0.17	3.22
lag(Y, 5)							0.10	1.52
lag(Y, 6)							–0.16	-4.03
Positive Market Movement: $(R_M>0)\cdot R_M$								
lagp(R_M, 0)							0.04	1.81
lagp(R_M, 1)	0.14	8.66	0.09	4.35	0.12	3.85	0.13	5.50
lagp(R_M, 2)			0.01	0.47	–0.02	-0.59	–0.03	-1.22
lagp(R_M, 3)			0.00	0.29	–0.01	-0.35	–0.02	-1.05
lagp(R_M, 4)							0.09	3.80
lagp(R_M, 5)							0.01	0.96
lagp(R_M, 6)							–0.01	-0.26
Negative Market Movement, Absolute Value: $(R_M<0)\cdot\lvert R_M\rvert$								
lagn(R_M, 0)							–0.02	-0.85
lagn(R_M, 1)	0.09	13.43	0.05	4.80	0.07	4.78	0.04	3.23
lagn(R_M, 2)			0.03	2.28	0.01	0.67	–0.01	-0.33
lagn(R_M, 3)			0.01	0.35	0.01	0.23	–0.02	-1.01
lagn(R_M, 4)							0.05	3.76
lagn(R_M, 5)							0.04	2.11
lagn(R_M, 6)							0.00	-0.24
$\bar{R}^2$:	48.2%		48.9%		56.9%		66.2%	
N:	563		563		135		135	

Explanations: The dependent variable Y in this time-series regression is the *percent* 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 percent change in the S&P 500. The "positive market movement" variables assign zero on days when the market declined. The "negative market movement" variables are analogous, except that they use the absolute value of R_M, as indicated in the header. The T-statistics are adjusted for heteroskedasticity and two lags as in Newey and West (1987).

Interpretation: 1-day lagged and 3-4 day-lagged large market movements are robust positive predictors of increases in the sum-total number of RH holdings. The 1-day spike is stronger on positive than on negative rates of return days, but it occurs in both. The 3-4 day spike is roughly symmetric.

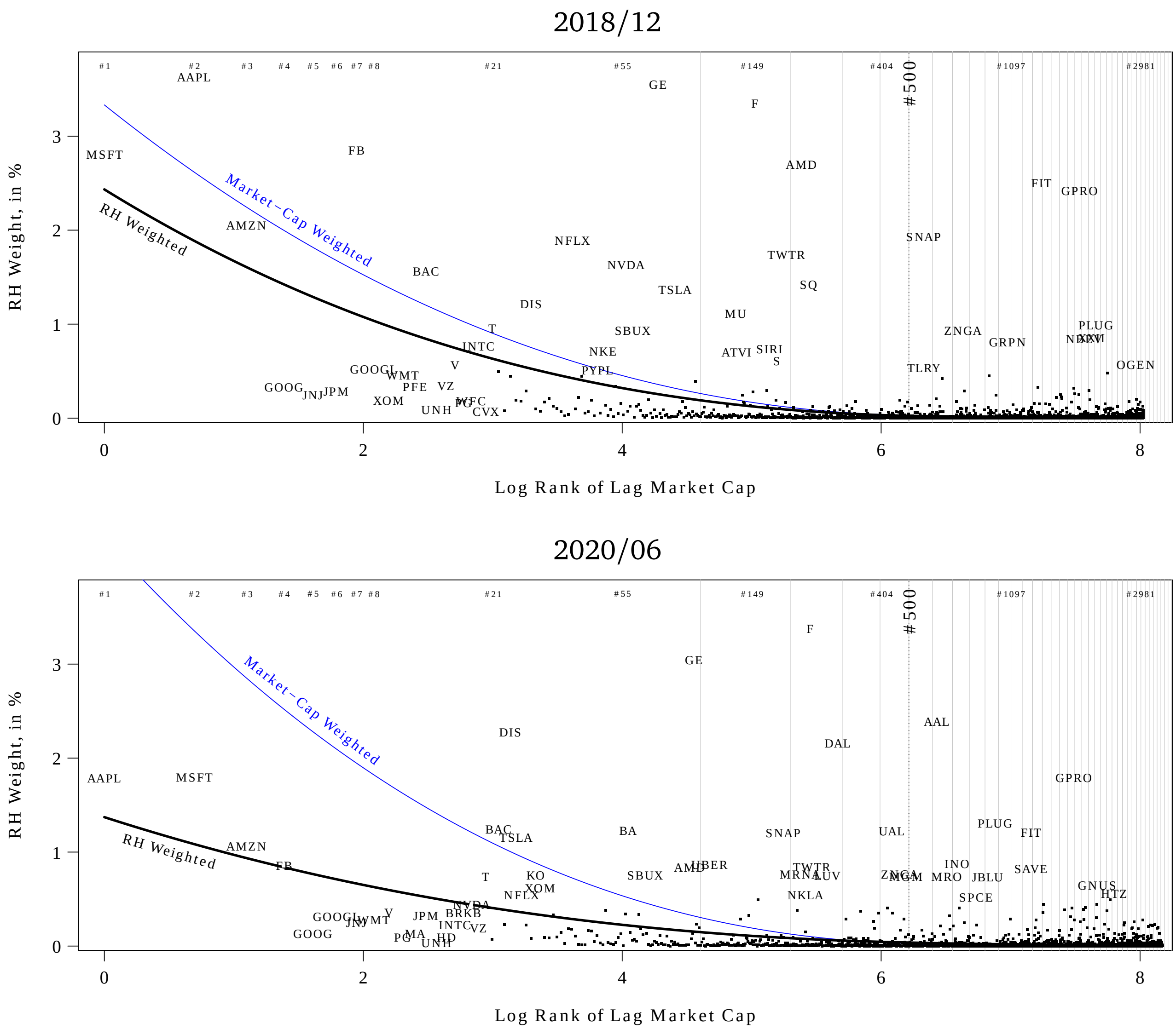
Table 4: Highest ARH vs VW Log-Investment-Rank Difference

Year	TIC	Business Description	Rank RH	Rank VW	ARH Pfio 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 Fin'l	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**

Explanations: RH is Robinhood, VW is value-weighted (always according to CRSP's shrout × |prc|). The stocks listed in this table (log-)ranked more highly in the investment holdings of RH investors than in the VW market portfolio either on 2019/12/31 or on 2020/06/30, such that they constituted at least 0.15% of the ARH investment weight. (All stocks listed in this table constituted less than 0.005% holdings in the VW portfolio.)

Source: 1-namelargeholds.R: 4 Sep 2020.

Figure 5: Snapshot ARH Weights by Log-Rank of Market Cap



Explanations: Each point is one stock's percentage investment weight in ARH against the log market-cap rank (from CRSP). For example, in 2019/12, AAPL was the biggest stock by marketcap, but it accounted for "only" 2.3% of the ARH portfolio. For stocks with unusually large weights and for the biggest stocks, the figures show ticker symbols instead of points. The black (loess) line fits the (RH) points. The blue line shows the line that a market-cap VW portfolio would have assigned (which would have been about 4% for AAPL).

Interpretation: Retail investors overweight many consumer-related tech stocks, as well as fallen angels (like Ford, GE, UAL) that probably remain because investors were reluctant to sell them. Retail investors systematically underinvest in the biggest 500 stocks, with notable consumer exceptions.

Source: 2-mkbyszplots.R: 4 Sep 2020 (files bysz-20*.pdf).

Table 5: ARH Investment Weights *Far Above* VW Investment Weights

		Pre-Covid				Post-Covid							
		Dec 2018				Jun2020				Holdings (in K)			Price
		Rank		Weight		Rank		Weight		Dec	Jun		Chg
Tic	Description	ARH	VW	ARH	VW	ARH	VW	ARH	VW	2018	2020	Chg	(RoR)
AAL	American Airlines	114	272	0.12	0.06	3	620	2.39	0.02	7	654	88.7	–0.64
AMD	AMD	6	218	2.70	0.08	21	92	0.84	0.20	161	229	0.4	1.92
BA	Boeing	53	26	0.29	0.78	11	57	1.23	0.34	17	337	18.5	–0.42
DAL	Delta Airlines	82	139	0.17	0.15	5	288	2.16	0.06	10	592	56.1	–0.47
DIS	Disney	17	27	1.22	0.69	4	23	2.28	0.66	72	624	7.6	0.04
F	Ford	3	152	3.35	0.13	1	233	3.38	0.08	200	925	3.6	–0.18
FB	Facebook	4	7	2.85	1.34	19	4	0.86	1.78	170	235	0.4	0.67
FIT	Fitbit	7	1388	2.51	0.00	13	1281	1.21	0.01	149	331	1.2	0.30
GE	General Electric	2	72	3.55	0.28	2	95	3.05	0.20	212	834	2.9	–0.05
GPRO	GoPro	8	1867	2.42	0.00	7	1785	1.79	0.00	144	491	2.4	0.19
INO	Vaccines	384	1931	0.03	0.00	17	725	0.88	0.01	1	241	133.4	5.41
JBLU	JetBlue Airlines	100	587	0.14	0.02	31	920	0.72	0.01	8	197	23.0	–0.36
LUV	Southwest Airlines	51	167	0.29	0.11	27	266	0.75	0.07	17	205	10.8	–0.33
MGM	MGM Casino-Entmt	180	301	0.07	0.05	30	488	0.74	0.03	4	202	44.3	–0.31
MRO	Marathon Oil	129	322	0.10	0.05	29	668	0.74	0.02	6	202	31.6	–0.55
MU	Micron Storage	18	131	1.11	0.15	97	99	0.19	0.19	66	53	–0.2	0.42
NBEV	Cannabis Drinks	24	1927	0.84	0.00	87	2637	0.23	0.00	50	61	0.2	–0.76
NFLX	Netflix	11	37	1.89	0.50	36	25	0.54	0.65	113	148	0.3	0.65
NVDA	Nvidia	13	56	1.63	0.35	42	17	0.44	0.76	97	121	0.2	2.13
PLUG	Hydrogen Fuel	19	2115	0.99	0.00	9	971	1.31	0.01	59	358	5.1	4.85
SAVE	Spirit Airlines	501	689	0.02	0.02	22	1282	0.82	0.01	1	225	174.0	–0.69
SNAP	Snapchat	10	559	1.93	0.02	14	189	1.21	0.09	115	330	1.9	2.66
SQ	Square Pymnts	15	230	1.41	0.07	61	145	0.33	0.12	84	89	0.1	0.68
TSLA	Tesla	16	82	1.37	0.24	15	24	1.16	0.65	81	317	2.9	1.73
TWTR	Twitter	12	194	1.74	0.09	20	236	0.84	0.08	104	231	1.2	–0.03
UAL	United Airlines	335	188	0.04	0.10	12	437	1.23	0.03	2	336	157.2	–0.60
XXII	Cannabis-Nctn	23	2050	0.85	0.00	103	2793	0.19	0.00	51	51	0.0	–0.67
ZNGA	Gaming (FB)	21	757	0.94	0.01	23	464	0.77	0.03	55	210	2.8	1.14
									All ARH:	5,973	27,390	3.6	0.24

Explanations: These are the companies with the largest (absolute) excess ARH weights over the value-weighted marketcap (VWM) weights either on 2018/12/31 or 2020/06/30. The holdings changes and rates of return in the last four columns are from 2018/12/31 to 2020/06/30.

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.

Source: 1-namelargeholds.R: 4 Sep 2020.

Table 6: Return Performance of the ARH Crowd Portfolio, 2018 – mid-2020

	Daily Positions			Monthly Positions		
	0-F	1-F	5-F	0-F	1-F	5-F
alpha	0.098	0.052	0.088	2.000	0.915	1.298
(T)	1.13	1.50	3.14	1.31	1.62	2.54
xmkt		1.13	1.00		1.31	1.03
smb			0.54			1.02
hml			0.20			–0.38
rmw			–0.02			0.16
cma			–0.71			0.15
$\bar{R}^2$		0.83	0.90		0.86	0.90
df	531	530	526	25	24	20

Explanations: These are alphas from factor regressions based on the implied performance of the ARH portfolio (formed on the basis of RH participation, as in equation 1). For example, if 200 RH traders held stock A and 100 RH traders held stock B, the ARH portfolio invests twice as much in A as in B (regardless of share price differences). After forming the ARH portfolio, I calculate the portfolio rates of return and evaluate its performance with respect to different benchmarks. The 0-F benchmark is the mean (net of the prevailing risk-free rate (about 0 in the 2018-2020 sample). The 1-F benchmark is the CAPM market-model. The 5-F benchmark is the Fama-French 5-factor model.

Interpretation: In a sample this short, only the daily evaluation seems reasonable. The ARH crowd portfolio performed quite well. It tilted towards small stocks that invested aggressively.

Table 7: Explaining ARH Investment Weights (in %)

Panel A: 2018-2019 Sample (N=1,308,437)

	const	Weight: Shr Vol	Weight: Dol Vol	Rank MCap: top500	Rank MCap: top100	Log-Days since IPO	$\bar{R}^2$, in %
	.... Kitchen Sink >50 Variables						66.60
(QRH)	–0.014	98.6	45.7				58.81
	–0.005	105.9	52.9	–0.0858			61.26
	–0.006	107.7	58.6	–0.0765	–0.099		61.86
	–17.0	108.3	58.0	–0.0691	–0.091	1.013	62.12

Panel B: 2020 Sample (N=417,933)

	const	Weight: Shr Vol	Weight: Dol Vol	Rank MCap: top500	Rank MCap: top100	Log-Days since IPO	$\bar{R}^2$, in %
	.... Kitchen Sink >50 Variables						61.87
(QRH)	–0.011	113.4	25.52				49.45
	–0.006	119.2	30.52	–0.0585			50.87
	–0.006	120.5	33.82	–0.0534	–0.0563		51.10
	–8.7	120.9	33.54	–0.0498	–0.0523	0.515	51.19

Panel C: Full Sample (N=1,726,370)

	const	Weight: Shr Vol	Weight: Dol Vol	Rank MCap: top500	Rank MCap: top100	Log-Days since IPO	$\bar{R}^2$, in %
all years	.... Kitchen Sink >50 Variables						63.99
(QRH)	–0.013	101.9	40.9				56.72
	–0.005	108.9	47.6	–0.0798			58.93
	–0.006	110.6	52.8	–0.0714	–0.090		59.45
	–15.0	111.2	52.3	–0.0649	–0.083	0.894	59.67

Explanations: This table shows the performance of regressions explaining the ARH investment weights in grand pooled regressions. $\bar{R}^2$ are based on residuals where fitted values were winsorized at 0 and renormalized to add to 1.

Interpretation: A model with only two variables—trading volume and dollar trading volume over the last year—can explain over 50% of the ARH investment weights. (Not shown, this is also the case if the prediction is purely cross-sectional on one specific day only.)

Source: 5a-qrh-predictweights.R: 4 Sep 2020.

Table 8: Implication of QRH Model Differences For QRH Investment Weights

Panel A: Correlation of Fitted Investment Weights based on 2018-9 vs 2020 Coefficient Estimates

	Dataset			
	2018-9	2020	All	2020/06/30
w_{qrh}	98.63%	98.24%	98.55%	98.00%

Panel B: Correlation of Fitted Investment Weights based on 2018-20 Coefficient Estimates, Kitchen Sink vs QRH Model

	Dataset			
	2018-9	2020	All	2020/06/30
w_{qrh}	96.02%	90.73%	95.07%	86.94%

Explanations: These are correlations between fitted investment weights implied by different regression models (and/or their coefficient estimates). All model-implied investment weights were truncated at 0 and renormalized to sum to 1.

Interpretation: Coefficient-estimate differences between the 2018-9 and the 2020 model for the QRH model seem unimportant. Coefficient model-estimate differences between the kitchen-sink model and the simpler two-model QRH are more concerning.

Source: 5d-cors: 12 Sep 2020.

Table 9: ARH and QRH Portfolio Return Correlations, 2018 – mid-2020

Panel A: Rate of Return Correlations (Over Time-Series in ARH Universe)

	QRH Portfolio Return		ARH Portfolio Return		
	IS	OOS	All Yrs	pre-2020	post-2020
QRH (IS)	*	99.99	97.91		
QRH (OOS)		*	97.82	96.68%	98.51%
ARH			*		

Panel B: Sample Slippage: ARH vs CRSP Set of Stocks, QRH Portfolio

	Adjustment	Correlation
QRH (RH Universe) vs QRH (CRSP Universe)	Raw	99.78%
	Net of Market	98.58%

Panel C: Net-of-Model Correlations

	Adjustment	Correlation
ARH vs. QRH (OOS, ARH Universe)	Net of Market	86.96%
ARH vs. QRH (OOS, CRSP-Universe)	Net of Market	83.49%
ARH vs. QRH (OOS)	Net of 5-F	79.01%
ARH vs. QRH (IS)	Net of 5-F	79.65%

Explanations: Unlike earlier tables (such as Table 8), this table investigates correlations not in investment-weight space but in rate-of-return space. The QRH portfolio weights are from the two-variable model in Table 7 (based on one year of share volume and dollar volume). The IS performance uses the estimated weight coefficients from the full sample, the OOS performance uses the coefficients from the post-2020 model with the pre-2020 volume data (and vice-versa). Panel B shows that there is some slippage when we include stocks on CRSP but not in the Robintrack data. Panel C shows that the correlation between ARH and QRH returns remains respectable even after subtracting ohter important factors.

Interpretation: The QRH investment strategy has return performance very similar to that of the ARH strategy. However, Panel C shows that a good part of this is due to common stock-market movements.

Source: 5cd-*.R: Sep 2020

Table 10: Return Performance of the QRH Portfolio, 2018 – mid-2020

Panel A: Daily Performance

	QRH Pfio Ret			Delta (to ARH, Table 6)		
	0-F	1-F	5-F	0-F	1-F	5-F
alpha	0.051	0.004	0.042	0.040	0.041	0.038
(T)	0.60	0.17	2.16	2.09	2.15	2.26
xmkt		1.14	1.05		–0.02	–0.05
smb			0.37			0.18
hml			0.27			–0.06
rmw			–0.05			0.03
cma			–0.28			–0.43
rsq		0.91	0.95		0.00	0.25
df	531	530	526	531	530	526

Panel B: Monthly Performance

	QRH Pfio Ret			Delta (to ARH, Table 6)		
	0-F	1-F	5-F	0-F	1-F	5-F
alpha	1.035	–0.097	0.505	0.824	0.868	0.642
(T)	0.67	–0.25	1.37	2.37	2.44	2.30
xmkt		1.37	1.14		–0.05	–0.11
smb			0.71			0.32
hml			–0.10			–0.28
rmw			0.29			–0.13
cma			0.34			–0.19
rsq		0.93	0.95		–0.01	0.45
df	25	24	20	25	24	20

Explanations: For methods, see Table 6. The QRH portfolio here was formed based on lagged one-year share volume and dollar share volume, in accordance with Table 7 in the "OOS" way (model coefficients from other part). The table here presents alphas and exposures from factor regressions based on the implied performance of the portfolio return time-series. The daily series allow daily portfolio readjustment; the monthly series do not.

Interpretation: The QRH portfolio performed reasonably well. Like the ARH portfolio, the QRH portfolio tilted towards small stocks with weak profitability. Although the QRH return time-series matched the ARH return time-series reasonably well in its factor exposures (except in its exposure to CMA, where it was less aggressive), it could not quite match the ARH portfolio's performance (alpha).

Source: 5b-compare-factors.R: 12 September 2020 *tbl:inspiredperf in rhperf-exhibits.tex, September 23, 2020*

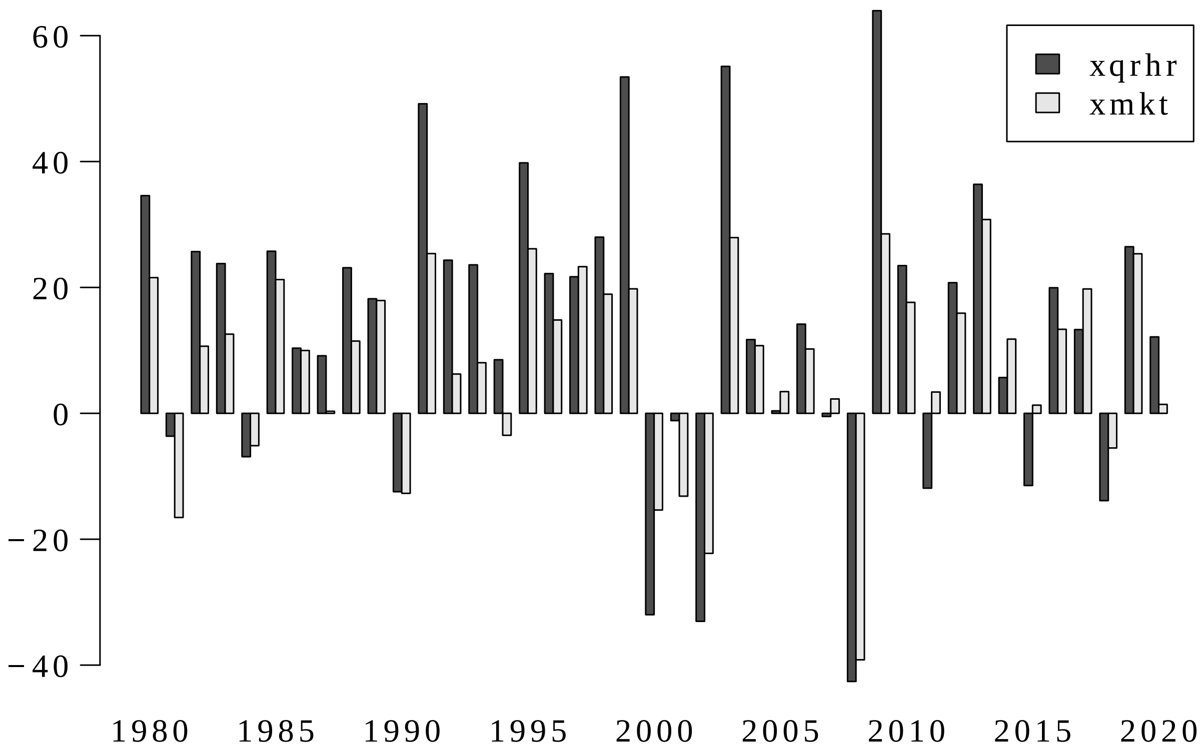
Table 11: Return Performance of the QRH Portfolio, 1980 – 2020

	Daily			Monthly			Annual		
	0-F	1-F	5-F	0-F	1-F	5-F	0-F	1-F	5-F
alpha	0.037	–0.005	0.014	0.76	–0.19	0.20	9.2	–1.9	1.6
(T)	2.53	–0.98	3.53	2.56	–1.59	1.77	2.68	–1.59	1.00
xmkt		1.26	1.16		1.45	1.20		1.32	1.16
smb			0.13			0.14			0.32
hml			0.13			0.05			–0.05
rmw			–0.58			–0.55			–0.43
cma			–0.51			–0.35			–0.04
$\bar{R}^2$		0.89	0.93		0.82	0.91		0.87	0.91
df	10,209	10,208	10,204	484	483	479	39	38	34

Explanations: For methods, see Table 6. This table is like Table 10, but the QRH strategy is extended to 40 years of data (from 1980 to 2020). (Not Shown: a delay of one day or month is largely inconsequential.)

Interpretation: The QRH portfolio would have underperformed on the 1-factor CAPM but outperformed on the Fama-French 5-factor model.

Figure 6: Return Performance of the QRH Portfolio After 1980



Explanations: These are returns from daily trading, then aggregated to yearly bases.

Interpretation: The performance of a QRH portfolio was different but not worse than that of the market. It was spectacularly good in 1999 and 2009, but even worse than the stock market in the already bad 2000.

Source: 6-ret-qrh-1980.R: 4 Sep 2020 (file qretail.pdf).