ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Economics and Business Economics

Robinhood Investors’ Exposure to Asset Pricing Factors

Name student: Leo Früngel

Student ID number: 526025

Supervisor: Dr. E. Smajlbegovic

Second assessor: Amar Soebhag

Date final version: 05/07/2022

Abstract

The rise in popularity of online trading apps has made the stock market increasingly accessible to retail investors. Robinhood has become a particularly prominent trading platform, notably due to the extreme risks and seemingly irrational trades made by its users. This paper studies the exposure of Robinhood traders to systematic factor trading strategies by constructing a Robinhood factor that goes long the most popular stocks and short the least popular ones traded on the app. I use ordinary least squares regressions to regress the returns of this factor on the returns of proven asset pricing factors, and find that Robinhood users are indeed exposed to all five of the asset pricing factors examined.

Table of contents

Introduction

Over the last decade, the stock market has become increasingly accessible to retail investors due to the emergence of online trading apps. These allow individuals to take direct control of their investments, making stock, option and cryptocurrency trading conveniently available at their fingertips. Among the most popular trading apps in the United States is Robinhood, which was founded in 2013 with the mission to "democratize finance for all" by making trading on its platform convenient and commission free. The app itself is designed to encourage user activity, and the company has been accused of 'gamifying' trading with features such as confetti showers on winning trades (a feature since disabled amid a lawsuit), with the average customer opening the app about seven times a day (Jakab, 2022). These characteristics helped it attract a younger and more inexperienced generation of traders during the highly volatile markets of the Covid-19 pandemic, leading the platform to have over 12 million users in 2020 and over 22 million users the following year (Statista, 2022). The bump in customers was also a result of the 'GameStop mania,' which saw small retail investors coordinate through social media (in particular Reddit) to take long positions in GameStop and squeeze hedge funds with sizeable short positions, leading the stock price to jump over 1600% in a 15-day span despite no positive change in the firm's underlying fundamentals.

Theoretical Framework

Investor Behaviour

Whilst GameStop was an extreme outlier, labelled by Robinhood's CEO as a "1 in 3.5 million event" (Rooney, 2021), it highlighted the importance of a field that has become of increasing interest in the world of finance; investor behaviour.

In their study of traders who switch from phone-based trading to online trading, Barber & Odean (2002) found that those who switch do so after a period of unusually strong performance, leading to overconfidence and self-selection into online trading. The overconfidence remains after switching, in part because of increased access to information which creates what the authors call the 'illusion of knowledge,' meaning that the accuracy of investors' forecasts improves much more slowly than the confidence in these forecasts.

Furthermore, the paper shows that traders who switch also suffer from an 'illusion of control,' which describes how the repeated and active involvement of individuals in their investments can lead them to believe they have control over the performance of their positions. These illusions lead investors who switch to trade more actively, more speculatively, and less profitably than before switching.

Much like Barber & Odean (2002), Grinblatt & Keloharju (2009) found that retail investors suffer from the overconfidence that they are better than the average investor (better-than-average effect), and from miscalibration (the confidence interval around their beliefs is tighter than the interval is in reality). They also identified that the number of trades made by individuals is related to sensation seeking, which is what Robinhood has been accused of attempting to elicit in its customers.

More recent research focused specifically on Robinhood traders shows that the app's users are more likely to engage in attention-induced trading than the average retail investor (Barber et al., 2021). Similarly, Welch (2020) found that Robinhood users are more likely to invest in salient stocks with above-average trading volumes.

Factor Investing

According to prominent asset pricing theories, none of the aforementioned heuristics or biases should play a role in determining investment strategies. Today, both academic research and financial institutions have pivoted towards investment strategies based on equity factors, which originate from Sharpe (1964), Treynor (1961), Lintner (1965) and Mossin's (1966) Capital Asset Pricing Model (CAPM). The CAPM assumes that markets are efficient, there are no taxes or transaction costs, investors are rational (they only hold efficient portfolios), investors can borrow or lend at the risk-free rate unrestricted, and that investors have homogenous expectations of market conditions. Based on these assumptions, the expected excess return of a security over the risk-free rate of return (R_i-R_f) is a linear function of the security's sensitivity to market risk, given by equation 1.1:

$$R_{it}-R_{ft} = a_i + B_i\,(R_{Mt} - R_{ft}) + e_{it} \tag{1.1}$$

$$B_i = Cov(R_i, R_M) / Var(R_M) \tag{1.2}$$

Where a_i is a constant and measures excess returns not captured by the model, B_i measures the volatility of a stock relative to that of the market as a whole (calculated using equation 1.2), R_M is the return on the market portfolio, and e_i is the error term.

Since the development of the CAPM, economists have attempted to find additional factors beyond market risk that can explain stock returns. Fama & French (1992) showed that a portfolio that is long small market capitalization stocks and short big market capitalization stocks produces positive risk-adjusted returns, resulting in the small minus big (SMB) or 'size' factor. In addition to this, they found that a portfolio that is long stocks with high book-to-market ratios and short stocks with low book-to-market ratios also produces positive risk-adjusted returns, resulting in the high minus low (HML) or 'value' factor. This led to the Fama-French three factor model:

$$R_{it}-R_{ft} = a_i + B_{1i} (R_{Mt} - R_{ft}) + B_{2i} SMB_t + B_{3i} HML_t + e_{it} \tag{1.3}$$

The Fama-French three factor model would later be expanded upon with a momentum (MOM) factor, which entails going long on stocks that performed well over the past 12 months and short on stocks that performed badly over the same period (Carhart, 1997). This led to the Fama-French-Carhart factor model:

$$R_{it}-R_{ft} = a_i + B_{1i} (R_{Mt} - R_{ft}) + B_{2i} SMB_t + B_{3i} HML_t + B_{4i} MOM_t + e_{it} \tag{1.4}$$

Recently, Fama & French (2015) would expand their previous three factor model to five factors, identifying the new profitability and investment factors. The profitability factor entails going long on stocks with robust profitability and short those with weak profitability (RMW), whereas the investment factor involves going long on stocks with conservative investment policies and short those with aggressive investment policies (CMA):

$$R_{it}-R_{ft} = a_i + B_{1i} (R_{Mt} - R_{ft}) + B_{2i} SMB_t + B_{3i} HML_t + B_{4i} RMW_t + B_{5i} CMA_t + e_{it} \tag{1.5}$$

Research Question and Hypotheses

Given the existing evidence that investing based on certain factors can lead to positive risk adjusted returns, combined with evidence that retail investors are irrational, one can question whether they would be exposed to prominent asset pricing factors. This leads me to the following research question:

Are Robinhood investors exposed to systematic factor trading strategies?

While extensive research has been done on the behaviour of individual investors, Robinhood investors specifically have not been studied in much depth; my research aims to add to the literature by combining the existing insights with knowledge on factor investing to determine whether there is a link between the two areas of study.

Based on research showing that the most popular Robinhood stocks exhibit negative returns over the next period (Barber et al., 2021), and taking factor strategies' positive risk-adjusted returns into account, one can hypothesize that Robinhood users are negatively exposed to proven asset pricing factors. I therefore define the following hypotheses to answer my research question:

Hypothesis 1: On average, Robinhood users are negatively exposed to the size factor (SMB) strategy

Hypothesis 2: On average, Robinhood users are negatively exposed to the value factor (HML) strategy

Hypothesis 3: On average, Robinhood users are negatively exposed to the momentum factor (MOM) strategy

Hypothesis 4: On average, Robinhood users are negatively exposed to the profitability factor (RMW) strategy

Hypothesis 5: On average, Robinhood users are negatively exposed to the investment factor (CMA) strategy

Data

Samples, variables and sources

My research focuses on the U.S. stock market and U.S. traders, as Robinhood usership is restricted exclusively to the United States. The sample period of analysis lies between 02/05/2018 and 13/08/2020 due to the availability of data on Robinhood users. To perform my analysis, I use data on Robinhood traders' stock preferences, the returns on these stocks, and daily returns on the equity factors described in the literature review: R_M-R_f (hereafter MKTRF), SMB, HML, MOM, RMW, and CMA.

I obtain the data on Robinhood trades from Robintrack (2022), which provides data obtained from Robinhood on the number of users holding each stock tradeable on the platform (exclusively U.S. stocks) between 02/05/2018 and 13/08/2020. Using this data, I construct a daily factor to represent users' stock preferences that is long the most popular stocks and short the least popular stocks traded on the platform, and will hereafter refer to this factor as the RH factor. To construct the RH factor, I first divide the stocks into daily portfolio quintiles based on the number of users holding them, using the earliest available daily observation (generally observed just after midnight). Then, I find both the equal-weighted and value-weighted (using the 1-day lagged market capitalization) return of each portfolio on each trading day. Finally, I find the difference in the returns between the highest and lowest quintile portfolios. I obtain the stock returns data from the Center for Research in Security Prices (CRSP) (2022), a database which provides daily and monthly price, return, and volume data on U.S. stocks with primary listings on the NYSE, AMEX and NASDAQ exchanges.

Throughout the process of gathering data on Robinhood users' holdings and stock returns, I remove 2,133,332 observations due to the absence of returns data on those days (these days were in the sample to begin with as Robintrack (2022) includes data on the number of users holding stocks on non-trading days). Furthermore, I remove 409,680 observations due to the absence of data on the number of users holding. Finally, I remove 7,233 observations due to duplicated daily observations which occur because of A and B class shares of firms. This results in a final sample of 3,760,124 observations across 8,074 stocks and 564 trading days.

I obtain the returns on daily equity factors from the Kenneth R. French Data Library (2022), which provides data on a variety of asset pricing factors based on the CRSP database. The factors are constructed based on stocks that are listed on the NYSE, AMEX or NASDAQ exchanges. As the sample period for the analysis is limited by the availability of Robintrack data, I again only use data from the period between 02/05/2018 and 13/08/2020.

Descriptive statistics

Table 1.1 shows descriptive statistics on the constructed RH factor, giving an overview of the stocks that are in the long (top quintile) and short (bottom quintile) portfolios of Robinhood users holding in terms of their trading volume, stock price, market capitalization (cap), price-to-earnings (P/E) ratio, debt-to-equity (D/E) ratio, quick ratio, and dividend yield (using data obtained from the CRSP (2022) database).

We can see from table 1.1 that the mean number of users holding in the top quintile is around 9,938, while the mean users holding in the bottom quintile is around 13. The difference in mean trading volumes between the quintiles is notable, with stocks the top quintile having an average volume of 3,889,868 and stocks in the bottom quintile having an average volume of around 58,018. This aligns with Welch's (2020) findings that Robinhood users prefer stocks with above average trading volumes. The difference in mean market cap between the quintiles is also notable, with an average of 19,000,000,000 dollars in the highest quintile and an average of 348,000,000 dollars in the lowest quintile.

The P/E ratio is one of the most commonly used measures of a stock's value, as it indicates how much an investor has to pay per dollar of the company's earnings. Therefore, a high P/E ratio generally indicates that investors expect high future earnings, whereas a low P/E ratio can indicate the firm in question has weak, or even no earnings at all (a characteristic of young firms and start-ups). The mean P/E ratio of stocks in the top quintile is 5.305, with the mean ratio of stocks in the bottom quintile being much higher at 22.640. However, this does not necessarily mean Robinhood investors simply prefer companies with lower earnings; companies with low P/E ratios can be considered undervalued, as their stock price does not yet reflect the potential future earnings that are only expected by few investors. As found by Barber & Odean (2002) and Grinblatt & Keloharju (2009), retail investors suffer from the overconfidence that they are better than the average investor, which could explain why Robinhood investors prefer holding stocks with lower P/E ratios; they believe that they know more about the future earnings of the supposedly undervalued companies than the average investor.

Another commonly used ratio used to analyse stocks is the D/E ratio, which indicates how much leverage a company is using. A firm with a high D/E ratio is often associated with more risk to shareholders, as it means the company is financing a significant part of itself through debt. The mean D/E ratio of 11.453 in the top quintile is much higher than the mean D/E ratio of 5.014 in the bottom quintile. This suggests that Robinhood users are willing to hold riskier stocks of less solvent companies, likely in the hope that the higher leverage will lead to significant investments in the companies' growth and thereby their future earnings.

The quick ratio is a measure of a company's liquidity by showing its ability to pay its current liabilities without selling its inventory. A quick ratio of 1 means the firm holds the exact amount of assets needed to pay its liabilities if liquidated (the higher the ratio, the more liquid the firm). The top and bottom quintiles have similar quick ratios on average, with the former being 2.663 and the latter being 3.023, suggesting that Robinhood investors are not particularly concerned with the liquidity of the companies they invest in.

The dividend yield is a measure of the direct pay out of companies to their investors. The dividend yields between the quintiles are similar, with the top quintile yielding a mean of 0.011 and the bottom quintile yielding a mean of 0.014. While the dividend yields do not greatly differ, the slightly lower yield in the top quintile could indicate that Robinhood users prefer stocks of younger firms that are less likely to pay out dividends than their more established counterparts.

Table 1.1: Descriptive statistics of the top and bottom quintiles of Robinhood users holding

Variable	Mean	SD	Min.	Max.
Top quintile				
Users Holding	9,937.902	34,536.690	568	986,629
Volume	3,889,868	10,000,000	0	1,000,000,000
Stock Price	46.235	99.809	0.074	3,225
Market Cap	19,000,000,000	66,200,000,000	30,780	1,970,000,000,000
P/E Ratio	5.305	70.445	-427.3125	481.067
D/E Ratio	11.453	519.770	-2,401.095	29,585
Quick Ratio	2.663	3.744	0.042	66.403
Dividend Yield	0.011	0.023	0	0.682
Bottom quintile				
Users Holding	12.702	9.317	0	54
Volume	58,017.880	1,011,153	0	491,000,000
Stock Price	40.154	109.746	0.038	4,699

Market Cap	348,000,000	1,460,000,000	63,540	83,100,000,000
P/E Ratio	22.640	60.506	-427.313	481.067
D/E Ratio	5.014	5.314	-169.425	16.852
Quick Ratio	3.023	8.017	0.003	225.169
Dividend Yield	0.014	0.015	0	0.256

Note. Users Holding, Volume, Stock Price and Market Cap are based on daily observations, whereas the P/E Ratio, D/E Ratio, Quick ratio and Dividend Yield are based on monthly observations. Volume describes the total number of shares sold on a given day. Market Cap (measured in U.S. dollars) is given by multiplying the total number of shares outstanding by the stock price on a given day. The P/E Ratio is calculated by dividing the stock price by earnings per share, and excludes extraordinary items. The D/E Ratio is calculated by dividing total liabilities by shareholder's equity. The Quick Ratio is calculated by dividing current assets net of inventories by current liabilities. The Dividend Yield is the dividend rate as a fraction of the stock price.

Table 1.2 shows descriptive statistics on the MKTRF, SMB, HML, MOM, RMW and CMA factors, as well as the constructed equal and value-weighted RH factors. From Table 1.2, we can see that the mean daily return is not positive for all of the examined factors over the period. On average, the market outperformed the risk-free rate of return by 0.053% daily, with the MOM and RMW factors also yielding positive daily returns at 0.014% and 0.012% respectively. The remaining factors (SMB, HML, and CMA) yielded negative daily returns on average, at -0.031%, -0.078%, and -0.009% respectively.

In addition to this, it is noteworthy that both the equal and value-weighted RH factors yielded positive daily returns on average, at 0.063% and 0.068% respectively. The difference between my findings and those by Barber et al. (2021) showing negative returns can likely be explained by the difference in the time horizons of the returns, as well as the method by which Robinhood trades are imitated. I use the RH factor which examines daily returns by using long and short portfolios, whereas Barber et al. (2021) showed that the top 0.5% of stocks bought by Robinhood users each day had an average monthly return of approximately -5%. In addition to this, my findings align with those of Welch (2020), who surprisingly finds that Robinhood users experienced positive daily returns when representing their trades with a portfolio that contains stocks weighted by the number of users holding (therefore without short-selling).

Table 1.2: Descriptive statistics of established equity pricing factors

Variable	Mean	SD	Min.	Max.
MKTRF	0.053	1.586	-12.00	9.34
SMB	-0.031	0.757	-4.52	5.73
HML	-0.078	1.044	-4.76	4.62
MOM	0.014	1.240	-6.11	5.93
RMW	0.012	0.421	-1.76	1.67
CMA	-0.009	0.381	-1.73	1.40
RH EW	0.063	0.993	-4.870	5.952
RH VW	0.068	0.614	-2.509	3.340
N	564*			

Note. All factor returns are based on daily observations and are given in percentages (%). The returns on MKTRF, SMB, HML, RMW and CMA are calculated by the 5 factor 2x3 sort, meaning that they are constructed using 6 value-weight portfolios formed on size and book-to-market, 6 value-weight portfolios formed on size and operating profitability, and 6 value-weight portfolios formed on size and investment (French, 2022). The returns on MOM are calculated using the intersections of 2 portfolios formed on size and 3 portfolios formed on prior (2-12 month) return. The daily size breakpoint is the median NYSE market equity. The daily prior (2-12 month) return breakpoints are the 30th and 70th NYSE percentiles. MOM is the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios, (French, 2022). RH EW and RH VW are the returns on the RH factor using equal and value-weighted returns, respectively. *N = 564 for all variables except RH VW, for which N = 563.

Methodology

I carry out my analysis using Ordinary Least Squares (OLS) regressions of the returns of the equal and value-weighted RH factors (dependent variables) on the returns of the equity factors described in hypotheses 1-5 (independent variables). I also use MKTRF as a control variable in all of the regressions to control for the relationship between excess market returns and users' returns. I apply both equal-weighted and value-weighted returns in my analysis, as both are commonly used academically and in financial practices. The most significant different between the weightings is that equal-weighted returns imply giving more weight to stocks with smaller market caps than value weighting would. This also means that an equal-weighted portfolio is exposed to higher risk than a value-weighted counterpart. If the relevant regression coefficient is significant at the 5% level, we observe a relationship between the RH factor and the given equity factor on the right-hand side of the regression equation. This results in ten regressions of the following (general) form:

$$RH = a + B_1x + B_2MKTRF + e_i \tag{2.1}$$

Where a is a constant and measures daily RH returns not explained by the equity factor, B_i are regression coefficients, x is the equity factor for a given hypothesis, and e_i is the error term.

For an OLS regression to yield the best linear unbiased estimator, it must exhibit constant residual variance (homoskedasticity), as well as uncorrelated error terms. Therefore, I first carry out a White test on each regression, with the null hypothesis of homoskedastic errors and an alternative hypothesis of heteroskedastic errors. If the p-value of the White test is smaller than 0.05, the null is rejected and heteroskedasticity-robust standard errors must be applied. To test for autocorrelated errors, I carry out a Breusch-Godfrey test on each regression, with the null hypothesis of uncorrelated errors and an alternative hypothesis of autocorrelated errors. If the p-value of the Breusch-Godfrey test is smaller than 0.05, the null is rejected and autocorrelation-robust standard errors must be applied. Newey-West standard errors are heteroskedasticity and autocorrelation consistent, and are therefore the appropriate standard error type in my regressions. A lag of 4 trading days is applied to the Newey-West standard errors (and in the Breusch-Godfrey test), as the general requirement for a consistent estimator is that the lag length be set to the integer of $T^{1/4}$, where T is the number of time periods in the sample (T = 564) (Newey & West, 1987).

In addition to this, to show the relationship between RH and all of the equity factors together, I build a regression model using the general to specific method (starting with many independent variables and removing insignificant ones until all regression coefficients are significant). I regress RH returns on the equity factor returns using an OLS regression with Newey-West standard errors (with a lag of 4 trading days), resulting in a multiple regression of the following form:

$$RH = a + B_1MKTRF + B_2SMB + B_3HML + B_4MOM + B_5RMW + B_6CMA + e_i \tag{2.2}$$

Results

The results of the White and Breusch-Godfrey tests can be found in tables 2.1 and 2.2 in the appendix. As expected, I found that that the use of Newey-West standard errors is warranted for all of the OLS regressions. For the regressions using the value-weighted RH factor (table 2.1), the five null hypotheses of the White test were all rejected, whereas the five null hypotheses of the Breusch-Godfrey tests were not rejected (at the 5% level). This means that the regressions exhibited heteroskedasticity, but they did not exhibit autocorrelated error terms. Still, the use of Newey-West standard errors is justified. As for the regressions using the equal-weighted RH factor (table 2.2), each of the five null hypotheses of the White and Breusch-Godfrey tests were rejected at the 5% level, meaning that all of the regressions exhibited both heteroskedasticity and autocorrelated error terms. The use of Newey-West standard errors is therefore also justified.

When analysing the results of my regressions, I noticed significant differences between the relevant regression coefficients depending on whether the RH returns were value or equal-weighted. A potential explanation for this difference is due to the mechanics of the weightings; equal-weighting weights small cap stocks more heavily than value-weighting. Therefore, given that retail investors' portfolios tend to be value-weighted (equal-weighting requires extremely frequent rebalancing), combined with the fact that the established equity factors were constructed using value-weighted portfolios (see notes of table 1.2), I draw my main interpretations and answer the hypotheses using the value-weighted returns. However, I also briefly discuss the results of the regressions using equal-weighted returns, as they may still be relevant and have implications on future research.

Regressions using value-weighted RH returns

The first column of table 2.3 shows the regression results relevant to the first hypothesis that, on average, Robinhood users are negatively exposed to the size factor (SMB) strategy. The regression coefficient of SMB is -0.342, and is significant at the 1% level, with a constant of 0.041%. This means that Robinhood users are negatively exposed to the size factor, as an increase in its returns is associated with a fall in RH returns. Therefore, when using value-weighted RH returns, one does not reject the first hypothesis.

A potential explanation for users' preference for large cap stocks (which the SMB factor goes short) coincides with the findings in my literature review, which showed that Robinhood traders buy stocks that are more salient (Welch (2020), Barber et al. (2021)). Firms with larger market caps are more likely to be 'household names' with exposure in the news, or whose products are visible and acquired by consumers on a daily basis, making them more likely to draw users' attention than stocks of smaller, lesser-known firms. A further potential explanation is that larger cap firms tend to release more information to the public about their operations and finances than their small cap counterparts, perhaps creating an 'illusion of knowledge' (Barber & Odean, 2002) within Robinhood traders with regards to bigger companies. My summary statistics also already indicated the negative exposure to size, as the top quintile portfolio of users holding had a notably higher mean market cap (19,000,000,000 dollars) than the bottom quintile portfolio (348,000,000 dollars).

The second column of table 2.3 shows the regression results for the second hypothesis: on average, Robinhood users are negatively exposed to the value factor (HML) strategy. The HML regression coefficient is -0.253, and is significant at the 1% level, with a constant of 0.032%. This means that Robinhood users are negatively exposed to the value factor, with an increase in its returns being associated with a drop in RH returns. Thus, when using value-weighted RH returns, one does not reject the second hypothesis.

Negative exposure to the HML factor indicates that on average, Robinhood users prefer growth stocks over value stocks. This can again be interpreted as a sign of overconfidence on the part of Robinhood traders, who prefer to buy stocks that have the potential for large gains over value stocks, which have proven to provide superior risk-adjusted returns in the long-run (Fama & French, 1992). The negative exposure may also be due the nature of the attention surrounding growth stocks, as these are often younger companies with innovative products and management teams which may be more appealing to the (on average) young Robinhood crowd than what could be perceived as 'stale' value companies.

The third column in table 2.3 shows the regression results for the third hypothesis that, on average, Robinhood users are negatively exposed to the momentum factor (MOM) strategy. The regression coefficient of MOM is 0.129, and is significant at the 1% level, with a constant of 0.052%. Thus, unlike the first two factors, Robinhood users are positively exposed to the momentum factor, with an increase in its returns being associated with an increase in RH returns. This means that when using value-weighted RH returns, one rejects the third hypothesis.

The positive exposure to momentum indicates that users follow the proven (Carhart, 1997) strategy of buying stocks that have recently seen an increase in price and selling stocks that have recently declined. That being said, users following a proven strategy does not necessarily mean that they are not biased, or being biased, in some manner. The Robinhood app allows users to easily access a “top movers” page just two taps away from the home page, which shows users the top 20 stocks with the largest daily gains and losses. While it is unlikely this is the main reason traders are positively exposed to a daily momentum strategy, it is plausible that the page contributes to their trades.

The fourth column in table 2.3 shows the regression results for the fourth hypothesis that, on average, Robinhood users are negatively exposed to the profitability factor (RMW) strategy. The RMW regression coefficient is -0.100, and is significant at the 5% level, with a constant of 0.055%. This means that Robinhood users are negatively exposed to the profitability factor, with an increase in its returns being associated with a fall in RH returns. Therefore, when using value-weighted RH returns, one does not reject the fourth hypothesis.

It seems especially irrational for traders to systematically expose themselves to unprofitable firms, as it may seem obvious that profitable firms will perform better than less profitable ones in the long run, even without taking Fama & French’s (2015) findings into account. A possible explanation ties into the negative exposure to the HML factor; users prefer growth stocks, which are often younger firms that are also less likely to be profitable. Robinhood traders could be overconfident in their ability to recognise which firms have the potential to shift their bottom lines from the red to the black.

The last column in table 2.3 shows the regression results relevant to the fifth hypothesis that, on average, Robinhood users are negatively exposed to the investment factor (CMA) strategy. The CMA regression coefficient is -0.410, and is significant at the 1% level, with a constant of 0.052%. Hence, Robinhood users are negatively exposed to the profitability factor, with an increase in its returns being associated with a fall in RH returns.

Consequently, when using value-weighted RH returns, one does not reject the fourth hypothesis.

The negative exposure could be due to aggressive investment policies being more likely to attract public attention. For example, if companies aggressively invest in research and development for new products, or if they decide to expand production or sales to a new geographical region, the investments could be misinterpreted by Robinhood users as positive indicators of the stocks' future outlook. However, in practice more aggressive spending is not a good indicator of positive returns, as shown by Fama & French (2015).

Table 2.3: OLS regression results for the relationship between value-weighted RH returns and equity factor returns

	RH VW				
Variable	(1)	(2)	(3)	(4)	(5)
SMB	-0.342**				
	(0.034)				
HML		-0.253**			
		(0.029)			
MOM			0.129**		
			(0.023)		
RMW				-0.100*	
				(0.058)	
CMA					-0.410**
					(0.069)
MKTRF	0.290**	0.296**	0.274**	0.260**	0.231**
	(0.026)	(0.016)	(0.028)	(0.023)	(0.017)
Constant	0.041*	0.032*	0.052**	0.055**	0.052**
	(0.017)	(0.016)	(0.019)	(0.020)	(0.018)
Observations	563	563	563	563	563

Note. The regressions are performed using Newey-west standard errors with a lag of 4 days, given in parentheses. The constants are the daily return on the RH factor not explained by the equity factor (given in percentages (%)). The first column regresses RH VW (value-weighted RH) on SMB and MKTRF, the second column regresses RH VW on HML and MKTRF, the third column regresses RH VW on MOM and MKTRF, the fourth column regresses RH VW on RMW and MKTRF, the fifth column regresses RH VW on CMA and MKTRF. *$p < 0.05$, **$p < 0.01$.

Regressions using equal-weighted RH returns

When using equal-weighted RH returns, I find that only 3 of the equity factors have a significant relationship with Robinhood trader preferences; size, momentum and profitability. The results of the relevant regressions can be found in table 2.4.

The SMB regression coefficient is 0.475, and is significant at the 1% level, indicating positive exposure to the size factor. While positive exposure would follow the findings of Fama & French (1992), I believe that one can attribute this finding to the mechanics of equal-weighting (giving far more weight to small cap stocks that value-weighting does), rather than to investor rationality. Given that the performance of small cap stocks naturally has an important effect on the SMB factor (which is long small cap stocks), the relatively heavier weight on small cap stocks could be the reason the returns of the SMB factor have a large and positive relationship with the returns of the equal-weighted RH factor, but not with the value-weighted one.

The regression coefficient for the MOM factor is -0.214 (significant at the 1% level), showing a negative exposure of Robinhood traders to the momentum factor. This would be interpreted as users holding stocks with declining prices and selling those with rising prices. The buying of stocks with falling prices is notable, as it coincides with the strategy of “buying the dip” which is commonly used by retail traders who are confident (at times overconfident) that the stock’s decline is only temporary and a rally will ensue.

Finally, under equal-weighting, Robinhood users have negative exposure to the profitability factor, with a regression coefficient of -0.248 (significant at the 5% level). This is the sole significant coefficient for which the sign (negative) aligns between the two weightings. Therefore, just as under value-weighting, a possible explanation for negative exposure to profitability could be that Robinhood users seem prefer growth stocks, and could be overconfident in their ability to recognise the firms that will successfully grow their currently low profits.

Table 2.4: OLS regression results for the relationship between equal-weighted RH returns and equity factor returns

	RH EW				
Variable	(1)	(2)	(3)	(4)	(5)
SMB	0.475**				
	(0.049)				
HML		0.091			
		(0.050)			
MOM			-0.214**		
			(0.039)		
RMW				-0.248*	
				(0.108)	
CMA					-0.308
					(0.185)
MKTRF	0.435**	0.467**	0.453**	0.488**	0.462**
	(0.034)	(0.037)	(0.036)	(0.038)	(0.044)
Constant	0.050	0.045	0.042	0.040	0.035
	(0.028)	(0.034)	(0.030)	(0.032)	(0.031)
Observations	564	564	564	564	564

Note. The regressions are performed using Newey-west standard errors with a lag of 4 days, given in parentheses. The constants are the daily return on the RH factor not explained by the equity factor (given in percentages (%)). The first column regresses RH EW (equal-weighted RH) on SMB and MKTRF, the second column regresses RH EW on HML and MKTRF, the third column regresses RH EW on MOM and MKTRF, the fourth column regresses RH EW on RMW and MKTRF, the fifth column regresses RH EW on CMA and MKTRF. $^{*}p < 0.05$, $^{**}p < 0.01$.

Regressing value-weighted RH returns on multiple factor returns

Table 2.5 shows the results of the regression of value-weighted RH returns on the returns of all of the established equity factors. The difference between the two models in columns 1 and 2 is the removal of independent variables with insignificant regression coefficients, namely the profitability (RMW) factor. Thereby, the model in column 2 shows that Robinhood traders are negatively exposed to the size, value, momentum, and investment factors (and are positively exposed to excess market returns). All of the signs and sizes of the coefficients are similar to those found in the simple linear regressions using value-weighted returns (table 2.3), except for the momentum factor whose coefficient becomes negative in the multiple regression.

Table 2.5: OLS regression results for the relationship between value-weighted RH returns and multiple equity factor returns

	RH VW	
Variable	(1)	(2)
SMB	-0.321**	-0.323**
	(0.032)	(0.032)
HML	-0.256**	-0.256**
	(0.039)	(0.039)
MOM	-0.149**	-0.151**
	(0.025)	(0.025)
RMW	0.021	
	(0.043)	
CMA	-0.230**	-0.226**
	(0.058)	(0.058)
MKTRF	0.293**	0.294**
	(0.016)	(0.016)
Constant	0.021	0.021
	(0.013)	(0.013)
Observations	563	563

Note. The regressions are performed using Newey-west standard errors with a lag of 4 days, given in parentheses. The constants are the daily return on the RH factor not explained by the equity factors (given in percentages (%)). The first column shows the full specification regression model, the second column shows the regression model restricted to variables with significant coefficients. $^{*}p < 0.05$, $^{**}p < 0.01$.

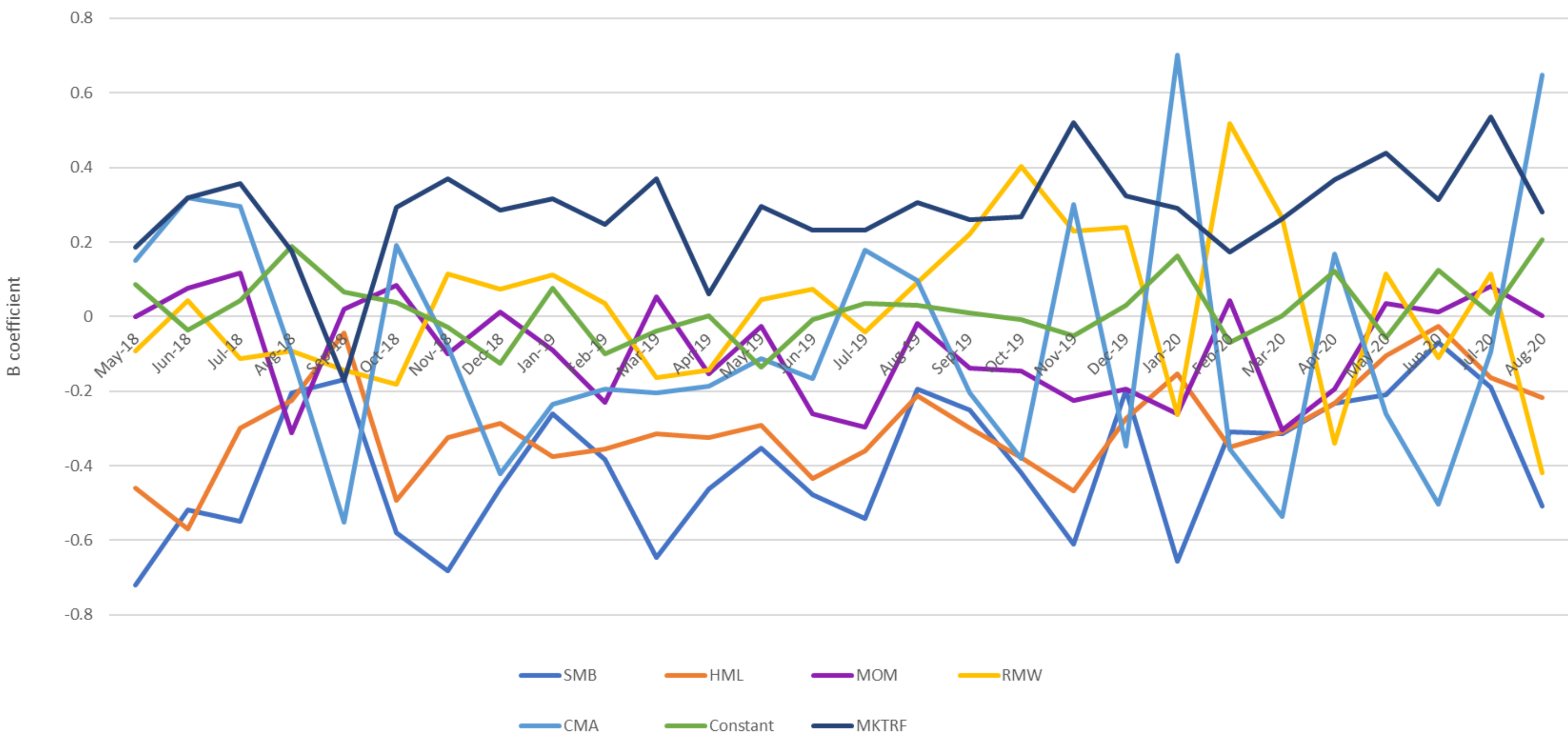
To see how the daily exposure to each factor develops over time, I also run the full specification regression from column 1 of table 2.5 as a monthly rolling window regression over the sample period, with the regression coefficients plotted in figure 1. From the figure, one can observe that there is no discernible trend in users' exposure to any of the factors, indicating that investors neither improved (positive trend) or worsened (negative trend) their investing practices over the period.

Users' exposure to both the size and value factors remained negative throughout the entire sample period, with the exposure to size being notably volatile. Exposure to the momentum factor fluctuated around 0, suggesting that users slightly preferred momentum strategies at times and 'buy the dip' or 'sell high' strategies at others. Exposure to the profitability factor was particularly volatile from the second half of 2019 until the end of the sample period, but should not be interpreted due to the insignificance of users' exposure to the factor shown in column 1 of table 2.5. Finally, the exposure to the investment factor was the most volatile, swinging frequently between positive and negative in a coefficient range of -0.6 to 0.7. While it is difficult to interpret what exactly could cause such significant swings over the period, a possible explanation could be that extreme herding behaviour drives the exposure to CMA; if the herd buys or shorts a particular stock or group of stocks with certain investment characteristics, it could swing the entire aggregate exposure of Robinhood users to the factor for a short amount of time. Users' exposure to market risk remains fairly constant and positive throughout, as does the constant of the regression.

Robustness check

Table 2.6 in the appendix shows the results of robustness checks in the form of the full specification regressions of value-weighted RH returns on the multiple equity factor returns (as in column 1 of table 2.5). I build the RH factor using various degrees (1 to 9 days) of lagged values of the number of Robinhood users holding to verify that the relationships between variables are not affected by the same-day buying pressure of users. The results show that no matter the degree to which user data is lagged, the same regression coefficients remain significant (with the same sign) as in column 1 of table 2.5 where no lag is applied. However, it should be noted that the magnitude of the coefficients and constants does vary slightly depending on the applied lag.

Figure 1: Monthly regression coefficients of the full specification rolling window regression of daily value-weighted RH returns on daily equity factor returns



Note. The regressions are performed using Newey-west standard errors with a lag of 2 months, following Newey & West (1987).

Conclusion & Discussion

After conducting my research using the constructed Robinhood factor, I can now answer my research question of whether Robinhood investors are exposed to systematic factor trading strategies. I find that Robinhood investors are indeed significantly exposed to all five of the studied equity factors when using value-weighted returns and controlling for excess market returns: negatively to size (SMB), value (HML), investment (CMA), and profitability (RMW), and positively to momentum (MOM). The exposures under equal-weighting are more ambiguous, but likely do not accurately reflect the behaviour of Robinhood users, as an equally-weighted portfolio requires perpetual monitoring and rebalancing.

While it is impossible to know exactly why each individual Robinhood investor holds a given stock, their negative aggregated exposures to multiple proven asset pricing factors (dependent on weighting) align with findings by Barber & Odean (2002), and Grinblatt & Keloharju (2009) that retail investors suffer from overconfidence and often make irrational trades. Users will invest contrary to academically and practically proven methods with the 'illusion' that their trades can outperform the standard.

I acknowledge that there are two principal limitations to my research. Firstly, my study was carried out using data spanning 564 trading days, which is a small sample period relative to existing research in the field of factor investing. This is reflected in the slightly negative mean daily returns of the SMB, HML, and CMA factors, which would on average have exhibited positive risk-adjusted returns if examined over a longer period. If new data regarding the trades of Robinhood users or other retail investors were to become available, it would provide an interesting avenue for future research to be conducted using a more extended time span, as well as an increased number of users trading on the platform. Improving the analysis with these elements could perhaps also reconcile the difference in results between the equal and value-weighted Robinhood factors. The second limitation is linked to the first; I analyse the returns of all factors at a daily frequency, but if additional data were to become available, it would be beneficial to repeat the analysis using monthly factor returns to examine whether the findings would differ.

Bibliography

Barber, B. M., Huang, X., Odean, T., & Schwarz, C. (2021). Attention-induced trading and returns: evidence from Robinhood users*. SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3715077

Barber, B. M., & Odean, T. (2002). Online investors: do the slow die first? *Review of Financial Studies, 15*(2), 455–488.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance, 52(*1), 57–82.

Center for Research in Security Prices, The University of Chicago Booth School of Business (2022). *CRSP Daily Stock.* Retrieved May 15, 2022.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance, 47*(2), 427–465.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics, 116*(1), 1–22.

French, K. R. (2022). *Kenneth R. French Data Library.* Retrieved May 17, 2022.

Grinblatt, M., & Keloharju, M. (2009). Sensation seeking, overconfidence, and trading activity. *The Journal of Finance, 64*(2), 549–578.

Jakab, S. (2022). *How Robinhood's psych experiment backfired horribly.* Retrieved from https://www.wired.com/story/robinhood-psych-experiment-gamestop/

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics, 47*(1), 13–37.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica, 34(*4), 768–783.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55(3), 703–708.

Primozic, C., & J., A. (2022). *Robintrack.* Retrieved from https://robintrack.net/

Rooney. (2021). *Here's what Robinhood executives allegedly said internally at the height of the gamestop short squeeze.* Retrieved from https://www.cnbc.com/2021/09/24/robinhood-what-executives-said-internally-during-gamestop-short-squeeze.html

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance, 19*(3), 425–442.

Statista. (2022). *Number of users of Robinhood 2014–2021.* https://www.statista.com/statistics/822176/number-of-users-robinhood/

Treynor, J. L. (1961). Market value, time, and risk. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2600356

Welch, I. (2020). The wisdom of the Robinhood crowd. https://doi.org/10.3386/w27866

Appendix

Table 2.1: P-values of the White tests for heteroskedasticity and Breusch-Godfrey tests for correlated errors on the regressions of value-weighted RH returns on equity factor returns

	RH VW				
Variable	(1)	(2)	(3)	(4)	(5)
SMB					
White Test	0.000*				
B-G Test	0.301				
HML					
White Test		0.000*			
B-G Test		0.919			
MOM					
White Test			0.000*		
B-G Test			0.891		
RMW					
White Test				0.000*	
B-G Test				0.914	
CMA					
White Test					0.000*
B-G Test					0.904

Note. The initial regressions are performed using default standard errors. The values shown are the p-values of the respective tests based on these regressions. The Breusch-Godfrey (B-G) tests are performed using a lag of 4. The first column regresses RH VW on SMB and MKTRF, the second column regresses RH VW on HML and MKTRF, the third column regresses RH VW on MOM and MKTRF, the fourth column regresses RH VW on RMW and MKTRF, the fifth column regresses RH VW on CMA and MKTRF. $*p < 0.05$.

Table 2.2: P-values of the White tests for heteroskedasticity and Breusch-Godfrey tests for correlated errors on the regressions of equal-weighted RH returns on equity factor returns

	RH EW				
Variable	(1)	(2)	(3)	(4)	(5)
<u>SMB</u>					
White Test	0.000*				
B-G Test	0.000*				
<u>HML</u>					
White Test		0.004*			
B-G Test		0.000*			
<u>MOM</u>					
White Test			0.003*		
B-G Test			0.000*		
<u>RMW</u>					
White Test				0.000*	
B-G Test				0.000*	
<u>CMA</u>					
White Test					0.000*
B-G Test					0.000*

Note. The initial regressions are performed using default standard errors. The values shown are the p-values of the respective tests based on these regressions. The Breusch-Godfrey (B-G) tests are performed using a lag of 4. The first column regresses RH EW on SMB and MKTRF, the second column regresses RH EW on HML and MKTRF, the third column regresses RH EW on MOM and MKTRF, the fourth column regresses RH EW on RMW and MKTRF, the fifth column regresses RH EW on CMA and MKTRF. *$p < 0.05$.

Table 2.6: Robustness check of the full specification regression of value-weighted RH returns and equity factor returns

	RH VW								
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SMB	-0.326*	-0.307*	-0.305*	-0.310*	-0.314*	-0.318*	-0.327*	-0.329*	-0.329*
	(0.033)	(0.034)	(0.033)	(0.032)	(0.033)	(0.034)	(0.035)	(0.035)	(0.034)
HML	-0.256*	-0.231*	-0.240*	-0.246*	-0.254*	-0.250*	-0.264*	-0.260*	-0.262*
	(0.040)	(0.043)	(0.043)	(0.041)	(0.041)	(0.044)	(0.046)	(0.045)	(0.045)
MOM	-0.156*	-0.109*	-0.112*	-0.107*	-0.115*	-0.115*	-0.123*	-0.121*	-0.116*
	(0.025)	(0.041)	(0.038)	(0.040)	(0.041)	(0.043)	(0.045)	(0.045)	(0.043)
RMW	0.000	0.079	0.076	0.087	0.083	0.088	0.097	0.098	0.103
	(0.048)	(0.062)	(0.060)	(0.060)	(0.062)	(0.061)	(0.061)	(0.060)	(0.059)
CMA	-0.236*	-0.280*	-0.271*	-0.259*	-0.259*	-0.258*	-0.250*	-0.259*	-0.253*
	(0.062)	(0.063)	(0.060)	(0.058)	(0.060)	(0.061)	(0.060)	(0.061)	(0.061)
MKTRF	0.282*	0.276*	0.274*	0.275*	0.271*	0.273*	0.275*	0.272*	0.273*
	(0.015)	(0.017)	(0.017)	(0.017)	(0.018)	(0.017)	(0.017)	(0.018)	(0.018)
Constant	0.019	0.014	0.009	0.011	0.013	0.012	0.009	0.009	0.008
	(0.013)	(0.013)	(0.014)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)

Note. The regressions are performed using Newey-west standard errors with a lag of 4 days, given in parentheses. The constants are the daily return on the RH factor not explained by the equity factors (given in percentages (%)). The columns show the full regression model using lagged observations of the number of users holding from 1 day of lag (column 1) to 9 days of lag (column 9). *p < 0.05.