Ville Valkama

Individual Investor and Volatility

Case Robinhood

School of Accounting and Finance

Master’s Thesis in Finance

Master’s Degree Programme in Finance

Vaasa 2024

VAASAN YLIOPISTO
Laskentatoimen ja rahoituksen yksikkö
Tekijä: Ville Valkama
Tutkielman nimi: Individual Investor and Volatility : Case Robinhood
Tutkinto: Kauppatieteiden maisteri
Oppiaine: Rahoitus
Työn ohjaaja: Timothy King
Valmistumisvuosi: 2024 **Sivumäärä:** 79

TIIVISTELMÄ:
Yksityissijoittajien merkitys osakemarkkinoilla on kasvanut viime vuosina. Tutkielmassa perehdytään osakekurssien volatiliteetin ja yksityissijoittajien omistuksen väliseen suhteeseen. Aikaisempi tutkimus on yhdistänyt yksityissijoittajat ja volatiliteetin laumakäyttäytymiseen, ei-informoitujen sijoittajien mukanaoloon (ns. noise trading) ja rationaalisista malleista poikkeaviin käyttäytymistapoihin. Yksityissijoittajiin liittyvissä volatiliteettia koskevissa tutkimuksissa on pääasiassa keskitytty aktiivisen kaupankäynnin tutkimiseen.

Tutkielmassa käytettiin ei-parametrisia testejä sekä lineaarisia regressioita datan analysointiin. Tutkimusaineistona hyödynnettiin Robintrack-aineistoa. Aineisto sisältää Robinhood-välittäjältä kerättyä tilastoa, joka mittaa osakkeiden suosiota käyttäjien keskuudessa. Tutkielmassa keskitytttiin tavallisiin osakkeisiin, joita otoksessa oli 3765 .

Kirjallisuudessa esitetään, että yksityissijoittajat saattavat olla kiinnostuneempia suuremman volatiiliteetin osakkeista. Toisaalta kirjallisuuden mukaan yksityissijoittajat ovat riskejä karttelevia, he saattavat systemaattisesti valita pienemmän volatileetin osakkeita. Ensiksi tutkielmassa tarkasteltiin yksittäissijoittajien omistuksen ja osakkeiden volatiliteetin samanaikaista riippuvuutta. Tavoitteena oli selvittää, onko yksityissijoittajilla yhteys suuremman volatileetin osakkeisiin. Toiseksi selvitettiin ovatko suuremman volatileetin osakkeet houkuttelevampia yksittäissijoittajille sekä seuraako yksityissijoittajien lisääntyneestä kiinnostuksesta osakkeisiin volatiliteetin kasvua.

Tulosten perusteella yksityissijoittajien omistamien osakkeiden ja niiden suuremman volatiliteetin välillä on yhteys. Suuremman volatiliteetin osakkeet näyttävät olevan houkuttelevampia yksityissijoittajille ja suurempi volatiliteetti voi johtua yksityissijoittajien kiinnostuksen lisääntymisestä.

Avainsanat: volatiliteetti, yksityissijoittaja, Robinhood

VAASAN YLIOPISTO
School of Accounting and Finance
Author: Ville Valkama
Thesis title: Individual Investor and Volatility : Case Robinhood
Degree: Master of Science in Economics and Business Administration
Programme: Master's Degree Programme in Finance
Supervisor: Timothy King
Year of graduation: 2024 **Number of pages:** 79

ABSTRACT:
This thesis explores the relationship between stock price volatility and individual investor ownership, specifically focusing on the users of the Robinhood app. Individual investors play an increasingly more important role in the financial markets. Previous research connects individual investors to volatility, primarily through noise trading, herding and deviations from the decision-making processes expected by the rational choice theory. Studies on volatility involving individual investors have predominantly centred on analysing retail trading.

The methodology consisted of non-parametric methods and linear regressions. We utilised the Robintrack dataset, which contains popularity metric data collected from the Robinhood brokerage, to study a sample of 3765 ordinary stocks. We first examined the contemporaneous relationship between individual investors and volatility. We sought to determine whether individual investors are associated with more volatile stocks. Based on the literature, we expected that individual investors may be associated with more volatile stocks; however, some risk-averse individual investors may systematically select lower-volatility stocks. Then, we examined whether individual investors are attracted to volatility and whether individual investor interest causes volatility.

Results indicate that individual investors are associated with more volatile stocks, even after controlling for company size. The positive relationship between volatility and individual investor holdings is consistent with the two stories. First, securities that exhibit higher volatilities may attract individual investors. Second, higher volatility may result from an increase in individual investor interest.

Keywords: volatility, individual investor, Robinhood

Contents

List of Figures

List of Tables

1 Introduction

Individual investors are often linked to volatility. Theoretical models suggest that individual investors, acting as noise traders, significantly affect volatility, and empirical studies have established a positive relationship between retail trading and volatility.

In recent years, increased retail investor participation in the financial markets has occurred. The growth in participation has been influenced by newcomers in the financial services industry and facilitated by removing barriers to entry for retail investors and barriers to trading. An example that highlights this trend is the substantial growth of an electronic trading platform called Robinhood.

The Robintrack dataset, which is publicly available, provides unique data that allows researchers to examine the behaviour of individual investors in the stock market. This dataset includes information on daily changes in retail investor ownership in securities available at Robinhood. The Robintrack dataset has yet to be utilised in prior studies to study the volatility of ordinary shares. Furthermore, financial literature on the relationship between investor ownership and volatility has typically focused on how institutional investor ownership affects and is affected by volatility.

Understanding the relationship between individual ownership and volatility is vital as individual investor participation in the markets increases. This thesis aims to address gaps in the literature by employing the Robintrack dataset to examine how volatility might influence or be impacted by individual investor ownership.

1.1 Background and motivation

The Robinhood app disrupted the brokerage industry by introducing zero-commission trades, eliminating account minimums, and providing a user-friendly app that integrated gamification elements (Eaton, Green, Roseman, & Wu, 2022; G. K. S. Tan, 2021). As a zero-commission brokerage, Robinhood generates revenue through premium services, securities lending, and payment for order flow, where user orders are forwarded to market makers for execution in return for a commission for the broker (Pagano, Sedunov, & Velthuis, 2021).

Robinhood users often exhibit traits similar to uninformed noise traders (Eaton, Green, Roseman, & Wu, 2021). Robinhood's strategy to acquire new customers can primarily explain this phenomenon. The company targeted inexperienced investors with incentives such as cash prizes for students to open accounts and a free stock lottery to encourage user referrals. As a result, Robinhood's user base skews younger and inexperienced, and, in particular, research indicates that the average Robinhood investor is 31 years old with modest balances (Moss, Naughton, & Wang, 2023). Marketing efforts led to a wave of new retail investors, and 3 million new accounts were reported in the first quarter of 2020 alone (Eaton et al., 2022; Pagano et al., 2021). By 2020, about five years after its launch, Robinhood had 13 million user accounts, and the brokerage was larger than established competitors like Schwab and E-Trade when measured by user volume (Eaton et al., 2022).

Robinhood trading positively correlates with retail trading (Baig, Blau, Butt, & Yasin, 2023; Eaton et al., 2021). Despite having smaller account sizes, Robinhood users have higher trading volumes than customers of other retail brokerages (Eaton et al., 2022). During 2020, retail investors contributed approximately 20% of stock market trading activity, and among the largest brokerage firms serving retail investors from Robinhood accounted for around 30% of daily trades (Barber, Huang, Odean, & Schwarz, 2022). The app's 'top movers list' feature highlights ten stocks with a market capitalisation of over $300 million that had the largest price changes since the previous day's close. According to Barber et al. (2022), this feature influences trading activity on the platform.

One aspect that needs to be mentioned that is absent from existing academic literature is Robinhood users' tendency to monitor stock prices frequently throughout the trading day. When the 'Robinhood Recap' feature launched in 2020, The New York Times reported that some users were startled by the frequency of their stock checks. (Marcus, 2020). For instance, one college student realised they had checked the price of one stock over 18,000 times in one year (Marcus, 2020). Thus, individual investor ownership of stocks could be primarily experienced through volatility.

Investigating the relationship between individual investors and volatility is a continuing concern within finance, as evidenced by the research from Kyröläinen (2008), Foucault, Sraer, and Thesmar (2011), Aharon, Baig, and Delisle (2022) and Baig et al. (2023). However, studies on analysing ownership have been limited to the institutional investor. A possible reason is that the research on changes in institutional ownership for U.S. stocks has been made more accessible by the requirement for institutional investment managers to file Form 13F quarterly with the Securities and Exchange Commission. On the other hand, micro-data related to individual investors' ownership of stocks has thus far been limited to proprietary sources. However, the Robintrack dataset makes this type of data openly accessible.

Sias (1996) discovered that institutional investors tend to be attracted to stocks with higher volatility, even when controlling for company size. This attraction to more volatile stocks goes against the predictions formulated from financial theories and even educated guesses. The study suggests that while the increase in volatility does not attract institutional investors, their actions may contribute to increased volatility. The Robintrack dataset provides an opportunity to replicate the methodology used by Sias (1996) with individual investor ownership data and enables an exploration of the relationship between individual investor ownership and intraday volatility.

1.2 Previous main studies

The Robintrack dataset has been prominently featured in the Journal of Finance by Barber et al. (2022) and Welch (2022). Using herding events, Barber et al. (2022) explored how attention-induced trading differed between Robinhood users and other retail traders. Welch (2022) examined Robinhood investors by evaluating the Robinhood consensus portfolio's performance using various models, including Fama and French (2015) five-factor model.

Researchers have also utilised the Robintrack dataset to explore many aspects of retail investor behaviours and their impacts, including volatility. Aharon et al. (2022) investigated the impact of Robinhood investor trading on the volatility of American depositary receipts. Eaton et al. (2021) identified the influence of Robinhood investors on market liquidity and return volatility during platform disruptions. Similarly, Eaton et al. (2022) investigated the differential effects of outages between Robinhood and traditional brokerage platforms. Meanwhile, Pagano et al. (2021) analysed changes in strategy among Robinhood traders during the COVID-19 pandemic to find out their impact on market quality. Ozik, Sadka, and Shen (2021) examined the impact of retail investors who traded stocks receiving high pandemic-related media coverage on stock liquidity. Bryzgalova, Pavlova, and Sikorskaya (2023) studied retail trading behaviours by examining the breadth of ownership and retail options trading on Robinhood. Finally, Moss et al. (2023) analysed individual investors' portfolio adjustments towards environmental, social, and governance press releases.

A growing body of literature recognises the relationship between individual investors and volatility. Studies by Kyröläinen (2008), Foucault et al. (2011), Aharon et al. (2022) and Baig et al. (2023) examined the relationship between trading of individual investors and volatility. Kyröläinen (2008) discovered a relationship between individual investor trading and increased volatility. Additionally, Kyröläinen (2008) identified that uninformed trades by individual investors contribute to volatility. Foucault et al. (2011) examined the French stock market during regulatory change. Their results highlight how actions by individual investors can cause volatility. Aharon et al. (2022) studied the impact of individual investors on the volatility of the American depositary receipts market. They showed that Robinhood users negatively affect the stability of the cross-listed securities. Baig et al. (2023) reported that during periods of market stress, the impact of retail investors on volatility is more substantial.

A considerable amount of literature has been published explaining how individual investor behaviour diverges from the decision-making processes expected by the rational choice theory. The behaviour of individual investors is impacted by what is known as the disposition effect (Shefrin & Statman, 1985), where they hold onto the belief that underperforming stocks will eventually outperform (Odean, 1998). Additionally, studies into individual investors' risk preferences have shown that while on aggregate, individual investors tend to favour stocks with characteristics resembling lotteries (Kumar, 2009), risk-averse investors prefer lower-volatility stocks (Dorn & Huberman, 2010).

Past studies have consistently shown that individual investors, often called noise traders, significantly impact asset price formation and stock volatility (Barber, Odean, & Zhu, 2009; Foucault et al., 2011; Nofsinger & Sias, 1999). In the seminal study about individual investor buying behaviour, Barber and Odean (2008) suggested that for individual investors to influence prices considerably, their trades must be substantially correlated as the size of their trades is generally small. They concluded that individual investors can influence prices in the direction of trades over shorter periods. On the other hand, individual investors have an asymmetric effect on daily stock volatility, as discovered by (Avramov, Chordia, & Goyal, 2006).

Similarly, Barber et al. (2022) found a difference in how Robinhood users influence stock prices compared to other retail investors. Their findings suggest that Robinhood users' heightened attention-driven buying behaviour increases trading and contributes to buy-side herding events, usually followed by negative returns. While individual investors are commonly associated with herding behaviour (Barber, Lee, Liu, & Odean, 2009), individual investors can also exhibit contrarian characteristics (Luo, Ravina, Sammon, & Viceira, 2022).

1.3 Purpose of the study and intended contribution

This thesis explores the relationship between individual investor ownership and volatility. Recently, there has been renewed interest in exploring how the growing number of individual investors impacts the stock market. What needs to be clarified is the nature of the impact of individual investor ownership directly on volatility. Additionally, previously published studies have utilised the Robintrack dataset to analyse various aspects of Robinhood users' trading habits and behaviours. However, they have not addressed how their ownership of ordinary stocks relates to volatility. This thesis seeks to address these gaps in the literature.

Information on institutional ownership is widely available as data can be compiled from regulatory filings. However, detailed data on individual investor ownership has been more scarce. Using the Robintrack dataset, we will follow the research design set out by Sias (1996) to explore how individual investor behaviour might affect or impact volatility. Additionally, by following the methodology found in Sias (1996), we can investigate whether patterns present in the relationship between institutional investor ownership and volatility exist with individual investors.

1.4 Hypothesis development

This thesis seeks to clarify this relationship by building upon existing research on how price fluctuations may influence or be influenced by individual investors. We pose three research questions to understand the connections between individual investor ownership, company size, and volatility. Specifically, we explore how volatility affects individual investor ownership and whether individual ownership levels can predict volatility.

First, what role does the interaction between company size and individual investor ownership play in shaping stock price volatility? Based on the findings by Fama and French (1993), Liew and Vassalou (2000), and Campbell and Vuolteenaho (2004), positive average returns of size-effect are attributable to greater risk of small firms. However, it is worth examining whether individual investor ownership impacts volatility even after controlling for company size. This research question will involve assessing research on individual investors' participation in the financial market and how risk-return trade-offs may affect their decision-making processes.

Second, how does a change in volatility influence individual investors' decisions about stock ownership? The literature presents conflicting evidence on whether individual investors are drawn to volatility or are volatility averse, i.e. risk averse. This research question will require evaluating the literature on individual investor behaviours and their relation to volatility.

Third, to what extent can past levels of individual investor ownership serve as an indicator of stock price volatility? The literature on noise trading suggests that individual investors can affect volatility. However, this effect may not be symmetrical. This inquiry will delve into theoretical and empirical literature to explore the impact of individual investors on financial markets.

We develop three hypotheses to test the relationships between individual investor ownership and volatility. Drawing from the research design in Sias (1996), we address the issue of reverse causality: whether heightened interest from individual investors precedes heightened volatility or vice versa. The second and third hypotheses account for this issue conjointly.

The first hypothesis explores, concurrently, the impact of company size, volatility and individual investor ownership:

Hypothesis H_0: Company size and the level of individual investor ownership do not significantly impact volatility.

Hypothesis H_1: Company size and the level of individual investor ownership significantly impact volatility.

The second hypothesis examines if and how individual investors might react to a change in volatility:

Hypothesis H_0: The change volatility does not significantly influence the level of individual investor ownership.

Hypothesis H_1: The change in volatility influences the level of individual investor ownership.

The third hypothesis aims to explore whether changes in individual investor interest are indicative of volatility:

Hypothesis H_0: Changes in individual investor ownership do not predict volatility.

Hypothesis H_1: Changes in individual investor ownership predict volatility.

1.5 Structure of the study

This thesis has five chapters. The first chapter is the introduction. The second chapter contains a literature review covering individual investor behaviour, specifically herding and contrarian investor behaviours. It explores the literature on the relationship between volatility and investor types and outlines the development of various volatility measures. The third chapter explains the data, measures and methodology used in the thesis. The methodology section explains the study's use of non-parametric methods and linear regressions. The fourth chapter presents the empirical results and discusses them in the context of past studies mentioned in the literature review. Finally, the fifth chapter concludes the study and suggests possible paths for future research.

2 Literature review

The existing literature offers differing perspectives on the relationship between volatility and individual investor behaviour. Some authors argue that risk-averse individual investors systematically select lower-volatility stocks as they are guided by heuristics that reflect their risk preferences. Others have suggested that individual investors favour stocks with high volatility and higher return skewness. Thus, the generalisability of published research regarding the relationship between individual investor ownership and volatility is problematic.

To date, several studies have highlighted factors associated with the connection between individual investor ownership and volatility. From them, we can formulate three assumptions. First, individual investors may favour small stocks, which can increase volatility as small stocks may have liquidity limitations. Second, a high level of individual investor ownership is generally associated with lower information quality and quantity. Third, as most academics argue, individual investors are likelier to behave irrationally and trade on 'noise' than institutional investors.

There are several reasons to suspect that securities favoured by individual investors may exhibit greater intraday volatility. Securities that exhibit greater volatility may attract individual investors. Individual investors, for example, may expect future returns to be higher for losing investments and stocks with higher volatility. Additionally, an increase in individual investor interest may induce volatility. Although individual investor trades are generally small, these trades may be correlated enough to impact prices. Thus, herding and contrarian trading may exacerbate price movements and increase volatility.

The following literature review offers a more detailed examination of various contexts related to the relationship between individual investors and volatility. This review will cover five major topics: behavioural finance, individual investor behaviour, herding and contrarian behaviour, investor types and volatility, and volatility measures.

2.1 Behavioural finance

Investors face a complex problem when selecting which stock to buy, as there are thousands of possibilities. Institutional investors tend to employ explicit purchase criteria that circumvent attention-driven buying (Barber & Odean, 2008). However, individual investors tend to display buying behaviour driven by attention (Barber & Odean, 2008). This difference emphasises the need to understand the factors that affect investment choices, a topic examined in behavioural finance.

Much of long-established economic and financial research hinges on the idea that humans, as rational agents, aim to maximise wealth and minimise risk. According to Fama (1970), the Efficient Market Hypothesis states that markets are efficient when they incorporate all available information. According to this theory, it becomes impossible for market participants to outperform the market consistently (Malkiel, 1989). Fama (1965) recognises that stock prices may be influenced by noise or intrinsic value.

The hypotheses presented in Chapter 1 do not fully align with market efficiency, as prices are expected not to reflect the efficient processing of information. Additionally, the research questions imply an underlying theme that factors beyond rational decision-making may influence individual investor behaviour. Empirical studies have highlighted that the behaviours of individual investors differ from what is expected based on the rational models. Critics of the Efficient Market Hypothesis argue that emotions and irrational behaviour influence price setting in financial markets. Shiller (2003) states that this leads to excessive volatility and elevated risk levels.

Behavioural finance is underpinned by the concepts of ’limits to arbitrage’ and ’psychology’ (Thaler & Barberis, 2002). ’Limits to arbitrage’ are based on noise trader risk and implementation costs. ’Psychology’ is supported by findings from empirical finance related to anomalies of twin shares and index inclusions. Various models, like those by De Long, Shleifer, Summers, and Waldmann (1990a) and Barberis, Shleifer, and Vishny

(1998), explain these concepts by addressing noise trader risk and investor reactions to new information, under- and overreactions due to investor overconfidence and the impact of slow information diffusion on market behaviour.

Much of the literature's theoretical explanations concerning the relationship between individual investors and volatility concern 'noise' in the financial markets. Black (1986) theorised that noise inflates the short-term volatility of prices, diverging from their actual value based on the interplay of information and noise. Consequently, daily price movements can have twice the variance of value movements until these variances eventually converge when the price returns closer to value over time. Noise traders significantly influence these fluctuations, notably favouring low-priced stocks, thereby causing enhanced volatility and liquidity. Such trading behaviour can distort pricing accuracy and obscure the relationship between price-to-earnings ratios and actual stock value, rendering it challenging to manage portfolios strategically despite the potential high returns.

De Long et al. (1990a) later expanded on this with a more formalised theoretical model of noise trading, which analysed the market in the context of risks emanating from unsophisticated investors and actions of rational arbitrageurs. They find that prices could deviate from fundamentals due to noise trader activities, even resulting in higher expected returns for such traders. The overlapping generations model reveals that arbitrage is subject to the risk of extreme misinterpretations by noise traders in future generations. As a result, rational arbitrageurs turn their attention to noise trading rather than fundamental information, interpreting the signals followed by noise traders to anticipate their actions.

Additionally to the above-described model by De Long et al. (1990a), De Long, Shleifer, Summers, and Waldmann (1990b) and Shleifer and Summers (1990) portray individual investors as noise traders. Furthermore, Eaton et al. (2021) and Friedman and Zeng (2022) find empirically that individual investors behave as noise traders. Contrary to previous studies, Kelley and Tetlock (2013) research findings challenge the noise trader hypothesis. They find that there is a systematic relationship between retail trading activity and future stock returns that cannot be explained by noise trader behaviour. This systematic relationship challenges the notion that noise traders dominate the market and suggests that retail traders may have some informational advantage or are reacting to meaningful signals in the market.

The theory related to noise trader risk, called synchronisation risk by Abreu and Brunnermeier (2002), disputes the claims behind the Efficient Market Hypothesis. It explains why behavioural biases in prices may persist even in the presence of rational arbitrageurs. The synchronisation risk is distinct from noise trader risk as behavioural traders do not directly cause it. Abreu and Brunnermeier (2002) posit that the rational arbitrageur may be uncertain about other arbitrageurs' actions to minimise holding costs. Thus, arbitrage trades are limited, and contrary to the Efficient Market Hypothesis, there can be considerable deviations from efficient prices. Baig et al. (2023) argued that synchronisation risk might be the most accurate explanation of the relationship between individual investor trading and volatility.

Many published studies describe the disposition effect as an explanation for individual investor behaviour that diverges from the expected behaviours in the rational models. The critical feature of prospect theory by Kahneman and Tversky (2013), or their approach to choice under uncertainty, is the aversion to loss realisation. As Grinblatt and Keloharju (2001) and Odean (1998) put forward, the disposition effect often leads to selling winning stocks too early and keeping losing stocks for too long, a phenomenon known as the disposition effect (Shefrin & Statman, 1985). Mental accounting, regret aversion, self-control, and tax considerations contribute to this behaviour. Although tax considerations alone do not account for these patterns, together with the three behaviours, they provide a more thorough explanation.

Models by Barberis et al. (1998), K. D. Daniel, Hirshleifer, and Subrahmanyam (1997), and Hong and Stein (1999) further illustrate investor reactions to new information. Barberis et al. (1998) present a case of potential overreactions to one-time news events triggered by Bayesian updating of beliefs based on regular earnings. K. D. Daniel et al. (1997), propose under-reactions and overreactions due to investors overestimating private data and undervaluing public signals. Finally, Hong and Stein (1999) report underreaction, momentum effects, and overreactions as consequences of limited rationality in processing all information simultaneously.

This section has reviewed the three key aspects of behavioural finance related to individual investors and volatility. These aspects form the foundation for addressing the three research questions. The theories related to 'noise' proposed by Black (1986), De Long et al. (1990a) and Abreu and Brunnermeier (2002) emphasise the influence of factors beyond rational information processing on volatility. Thus, if Robinhood investors' buying and selling of stocks is influenced by noise, and mispricing persists due to delayed arbitrage, changes in individual investor ownership should impact volatility. The disposition effect by Shefrin and Statman (1985) underscores how investor behaviour driven by past performance and perceived risk diverges from expected behaviours in rational models. The disposition effect further explains how changes in individual investor ownership could impact volatility. Furthermore, the models proposed by Barberis et al. (1998), K. D. Daniel et al. (1997), and Hong and Stein (1999) provide context for how volatility might influence individual investor ownership or be impacted by it.

2.2 Individual investor behaviour

There is a large volume of published studies describing individual investor behaviour. Literature can be treated under three general headings: performance of individual investors, individual investors' risk preferences, and individual investor risk management. We can gain insights into individual investor ownership by examining empirical literature on individual investor behaviours related to buying and selling. On the other hand, studying individual investors' risk preferences can explain whether individual investors are attracted to volatility. Finally, exploring literature on how individual investors manage risks will give us insights into how individuals act on and react to risk.

Individual investors' stock ownership depends on their performance, and the performance of individual investors has been a subject of interest in financial research for a considerable amount of time. Schlarbaum, Lewellen, and Lease (1978a) studied the stock performance of investors at a full-service brokerage firm, while Schlarbaum, Lewellen, and Lease (1978b) and Odean (1999) examined the profitability of common stock trades by individual investors. The first comprehensive study of the overall performance of self-managed individual investors was done by Barber and Odean (2000). Barber, Odean, and Zhu (2008) show that individual investors' buying and selling behaviours, observed through small trades, are highly correlated over time. These retail trade patterns forecast future returns in the short term, where trades impact stock prices, and in the long term, where small stocks are most affected.

Similarly, Kumar and Lee (2006) found that individual investors' buying and selling behaviour has a common directional component, and correlations exist in return comovements. Additionally, as Nofsinger and Sias (1999) and Kumar and Lee (2006) found, individual investors have more influence in stocks with small capitalisation. Furthermore, Kelley and Tetlock (2013) discovered that daily buy-sell imbalances from retail orders usually indicate monthly stock returns, with trends persisting for up to a year. Indications are accurate for stocks with more retail traders, suggesting that biases in the sample do not explain the result.

Barber and Odean (2000) highlight the underperformance of households with high stock trading activity. Grinblatt and Keloharju (2001), and later Barber and Odean (2013), find that transaction costs partially explain poor investing performance. In addition, it has been documented that individual investors have poor stock selection abilities (Barber & Odean, 2013). Furthermore, Brandt, Brav, Graham, and Kumar (2010) show higher idiosyncratic volatility levels in stocks held primarily by retail investors. They also find that events likely to attract the attention of retail investors lead to changes in volatility, which is consistent with the results of Barber et al. (2008).

As described in the previous section, the disposition effect and tax-loss selling significantly influence the retail investor's decision to sell a stock, as shown by Grinblatt and

Keloharju (2001). As Odean (1998) explains, investors may choose to keep losing investments and sell winning ones because they expect that current losers will perform better than winners in the future. These beliefs could be considered rational if the expected future returns for losing investments are higher than those for winners (Odean, 1998). On the other hand, the beliefs are irrational if individual investors persist in holding losing investments despite evidence showing that winners have higher expected returns (Odean, 1998).

Many investors maintain concentrated portfolios despite portfolio theory advocating for diversified holdings. As Dorn and Huberman (2005) describe, investors often hold no more than three stocks even when affordable mutual funds offer an easy diversification option. Barber et al. (2022) calculate, based on 2020 statistics, that Robinhood users have, on average, about three stock positions. Thus, in this case, Robinhood users represent the underdiversified investors described by Dorn and Huberman (2005).

Diversification, although beneficial, limits the potential for high returns. Investors with a strong desire for upside potential may forego diversification in pursuit of a chance for substantial gains. Mitton and Vorkink (2007) suggest that skewness-diversification trade-offs affect portfolio formation and asset prices. Theoretical modelling shows that a preference for skewness can lead to a lack of diversification. Empirical evidence indicates that underdiversified investors experience higher skewness in returns. Thus, investors intentionally select stocks to enhance skewness and prefer higher return skewness over traditional mean-variance portfolio construction. Balasubramaniam, Campbell, Ramadorai, and Ranish (2023) explain that lack of diversification with investors' preferences for characteristics such as market capitalisation, turnover, beta, book-to-market, volatility, and skewness significantly influence investors' choices. These characteristics create similar investor clusters. In addition, individual investors are susceptible to influence from geographical location, traditional media, and social media, as Barber and Odean (2013) and Eaton et al. (2022) suggest.

Dorn and Huberman (2005) argue that many investors adopt a sequential stock-picking approach rather than optimising their portfolio in totality, as suggested in modern portfolio theory by Markowitz (1952). According to Dorn and Huberman (2005), this behaviour is caused by difficulties in fully optimising portfolios and a tendency to focus narrowly on individual choices rather than considering their combined impact. Dorn and Huberman (2010) infer that while investors might neglect optimal diversification strategies, their choices are not random; risk-averse investors systematically select lower-volatility stocks, and portfolio adjustments often lead to improved diversification when new funds are contributed. Thus, individual investors rely on heuristics that reflect their risk preferences and may only sometimes follow the optimal portfolio construction principles.

One possible explanation of why individual investors may be attracted to more volatile stocks could be the propensity to gamble. The study by Kumar (2009) indicates that on an aggregate level, individual investors tend to favour stocks with characteristics resembling lotteries. Low-priced stocks exhibiting high skewness and volatility are likelier to be perceived as lottery-type investments. Along the same lines, Weiss-Cohen et al. (2024) identified two distinct behavioural patterns among gamblers. During low volatility environments, gamblers exhibited more investment-like behaviours with fewer trades, while during high volatility, their actions were more akin to gambling, characterised by increased trading activity.

The prevalence of high-risk, high-volatility investment options on trading platforms may give gamblers an alternative means to experience similar sensations as gambling by engaging in excessive trading (Kumar, 2009). However, as Kostopoulos, Meyer, and Uhr (2022) study finds, during periods of increased ambiguity, investors tend to exhibit less risk-taking behaviour, which persists over time. Furthermore, The data supports that sentiment effects are amplified when ambiguity is high. Kostopoulos et al. (2022) also show that more ambiguity-averse investors react strongly to ambiguity increases by reducing their exposure to the stock market.

These studies indicate that several behavioural factors affect the relationship between individual investors and volatility. Empirical studies provide answers to the three research questions, which can be summarised in four points. First, the combination of the tendency of individual investors to hold under-diversified portfolios and to engage in active trading may contribute to significant daily changes in individual investor holdings that are statistically observable using a daily time frame. Second, the preference to sell winning investments and the reluctance to sell losing investments suggests that a stock's past performance and perceived risk may influence levels of individual investor ownership. Third, individual investors have poor stock selection abilities and are susceptible to local bias and media influence. This combination could cause volatility or ownership of stocks to be influenced by volatility. Fourth, individual investors are attracted to more volatile stocks and may seek sensations similar to gambling in the stock market.

2.3 Herding and contrarian behaviour

Herd behaviour, or herding, is often a critical factor in periods of high volatility and market instability (Spyrou, 2013). This behaviour is observed among institutional and individual investors but manifests differently in the investor groups. Herding includes theoretical and methodical approaches that are diverse and disconnected (Raafat, Chater, & Frith, 2009). For example, contrarian strategies can be the antithesis of herding and momentum strategies as suggested by Kendall (2023); meanwhile, Luo et al. (2022) find that contrarian behaviour can contribute to momentum.

In financial literature, herding has a variety of meanings that range from following a trend (Nofsinger & Sias, 1999), excessive agreement (De Bondt & Forbes, 1999), convergent behaviour (Hirshleifer & Hong Teoh, 2003), correlated behaviour (Hwang & Salmon, 2004) to mutual imitation (Welch, 2000). In cognitive psychology, Raafat et al. (2009) define herding as the process of aligning individual thoughts or actions within a group due to local interactions without centralised coordination. In finance, the herd is framed either narrowly or broadly. The herd can include just a group of market participants or investors who trade in the same direction or follow each other into and out of the same securities at the same time (Nofsinger & Sias, 1999; Sias, 2004) or it can be contagious behaviour affecting the market as a whole (Akerlof & Shiller, 2009; Keynes, 1936; Shiller, 2002, 2015).

Raafat et al. (2009) divide various herding models into two separate yet interconnected perspectives: pattern-based and transmission-based approaches. Pattern-based approaches focus on the unchanging structure and relationships within a system, whereas transmission-based approaches focus on how information is shared. These transmission theories are further classified as rational versus emotional, automatic versus controlled, and conscious versus unconscious. Non-mentalising transmission-based approaches range from emotion to social contagion to priming and do not postulate mentalising as a critical transmission aspect. Mentalising transmission-based approaches focus on rational processes, where an individual consciously and deliberately evaluates others' information and signals.

The mentalising transmission-based approach is the most common way to model herding in finance. Herding is usually described as the process where individually rational people, influenced by others' choices, can collectively make irrational decisions – known as 'informational cascades', that do not align with the group's preferences (Raafat et al., 2009). Decision-makers might disregard their information and copy others' choices, thinking previous decision-makers have valuable details. Disregarding information can lead to a lack of information sharing, vulnerability to minor disruptions, and the emergence of market trends (Banerjee, 1992; Bikhchandani, Hirshleifer, & Welch, 1992; Hirshleifer & Hong Teoh, 2003; Welch, 2000).

Some writers (Avery & Zemsky, 1998; Cipriani & Guarino, 2005; Drehmann, Oechssler, & Roider, 2005) have researched the commonness of herding in relation to informational complexity. They found that herding is less common in markets without trading barriers when trading based on information, and informational cascades are less likely when market prices are flexible. Herding can influence asset prices in markets with complex information and uncertainty, whereas herding behaviour is less common in financial markets with simple information and pricing structures. Nevertheless, Barber et al. (2022) find it somewhat contrary in Robinhood's case. Robinhood can be characterised by simplified information presentation and a less experienced user base. Barber et al. (2022) suggest that this leads to impulsive buying decisions and concentrated trading on specific stocks, exacerbating herd behaviour among Robinhood users. Such herding events typically lead to negative returns. Although, as Cole (2014) highlights, the statistical measure of volatility remains indifferent towards price direction as spikes in volatility frequently coincide with price declines as prices drop faster than they rise.

Herd behaviour can be measured in two ways in empirical studies (Spyrou, 2013). One way is to examine specific investor types using micro or proprietary data. Another way is by studying aggregate price and market activity data to understand herding towards market consensus. Some commonly used measures were developed by Lakonishok, Shleifer, and Vishny (1994) and Sias (2004) for specific investors and by Christie and Huang (1995) and Chang, Cheng, and Khorana (2000) for market activity data.

Institutional investors such as pension fund managers and US mutual funds do not typically participate in herding or positive feedback trading (Grinblatt, Titman, & Wermers, 1995; Lakonishok, Shleifer, & Vishny, 1992; Nofsinger & Sias, 1999; Wermers, 1999). Other studies have concluded that small stocks and growth-oriented funds may engage in positive feedback trading, or these markets can have a significant price impact from herding. Institutions tend to follow each other in trading securities, often driven by information inference, as Sias (2004) shows. This behaviour is more pronounced in smaller and more volatile industries in the US (Choi & Sias, 2009). It impacts market stability with buy herds destabilising prices and sell herds stabilising them (Gutierrez Jr & Kelley, 2008).

Individual investors often display more herding behaviour than institutional investors (Kim & Wei, 2002). Herding behaviour can be attributed to several factors that align with findings from the earlier section on individual investor behaviour. Li, Rhee, and Wang (2017) find that individual investors tend to distribute their investments evenly across stocks, which suggests less informed decision-making. Additionally, according to Li et al. (2017), they heavily rely on public information and are susceptible to market sentiment and attention-grabbing events. Moreover, their herding behaviour is often evident through frequent small trades, as indicated by Barber, Lee, et al. (2009); L. Tan, Chiang, Mason, and Nelling (2008)

Methods often used to study herding behaviour at the aggregate level include those described by Christie and Huang (1995) and Chang et al. (2000). Findings from US data show no evidence of herding during significant price swings. However, Chang et al. (2000) discovered herding in Japan, South Korea, and Taiwan but not in the US and Hong Kong. They determined that macroeconomic factors influence investor behaviour in herding markets more than individual company information. On a similar note, Hwang and Salmon (2004) examined herding in US and South Korean equity markets, finding that market crises reduce the prevalence of herding. Additionally, L. Tan et al. (2008) documented instances of herding among both domestic and foreign investors in dual-listed Chinese stocks. Herding in ETF investors has been analysed separately. ETF investors do not exhibit herding behaviour during extreme market movements, except in sector ETFs (Gleason, Mathur, & Peterson, 2004).

Avramov et al. (2006) explored how the trading activity of contrarian and herding investors impacts the relationship between daily volatility and previous returns. They found that informed trades - linked to contrarian traders - tend to reduce volatility after a stock price increase, while non-informative trades - associated with herding investors - increase volatility after a stock price drop. These actions are consistent with rational expectation models. Contrarian trading involves selling when there are unexpected positive returns, while herding involves following the market trends and making trades based on the prevailing sentiment. Thus, these two distinct behaviours - herding behaviour's crowd-following and contrarian behaviour's individualistic approach impact volatility.

Generally, individual investors offer liquidity, react promptly to overnight returns and short-term news and employ both momentum and contrarian strategies, significantly impacting financial market quality during turbulent market conditions (Eaton et al., 2022; Grinblatt & Keloharju, 2000; Pagano et al., 2021). While retail investors show a disposition effect, it does not entirely explain their contrarian trading behaviour (Luo et al., 2022). Retail net inflows often oppose price momentum, especially in stocks with significant earnings surprises. This retail trader activity can contribute to price momentum by causing the market to underreact news, particularly in highly retail-traded stocks.

These studies provide important insights into the three research questions presented in Chapter 1. Herd behaviour, characterised by individuals following others' choices instead of their information, suggests that aggregate investor interest can significantly impact volatility and lead to periods of instability in the markets. Additionally, institutional and individual investors exhibit distinct herding behaviours. Institutions often follow each other in trading securities, particularly in smaller and more volatile industries. This behaviour can influence market stability, with buy herds destabilising prices and sell herds stabilising them. On the other hand, individual investors react promptly to short-term news and employ momentum and contrarian strategies, influencing market quality during turbulent conditions. Regarding individual investors' impact on volatility, contrarian traders tend to reduce volatility after stock price increases while herding investors increase volatility after price drops. Thus, when testing the hypothesis claims, returns should be examined alongside volatility and individual investor ownership.

2.4 Investor types and volatility

Much of the contemporary volatility literature examines either institutional or individual investors. The prevalent approach in many studies revolves around a classification scheme derived from noise trader models, which delineates between sophisticated investors—typically institutional investors—and noise traders, predominantly represented by individual investors.

The thesis' hypotheses establish a direct correlation between particular investor categories and volatility levels. The first hypothesis indicates that a company's market capitalisation and the level of individual investor ownership exert a notable impact on volatility. Additionally, the second hypothesis posits that changes in volatility can, in turn, affect the prevailing level of individual investor ownership. Finally, the final hypothesis suggests that changes in individual investor ownership can indicate the level of volatility.

The study by Sias (1996) about the relationship between institutional investors and volatility directly concerns two hypotheses made in the thesis. Sias (1996) proposes two hypotheses to explain this relationship. The first hypothesis posits that securities with greater volatility attract institutional investors. The second hypothesis poses that increased institutional holdings lead to increased volatility. Sias (1996) finds that institutional investor interest may induce volatility. Contrary to established theories in finance, the study finds that institutional investors are associated with higher levels of volatility when capitalisation is held constant. However, the study finds no evidence supporting the first hypothesis that securities with increasing volatility attract institutional investors, as results suggest that institutional investors are more likely to sell stocks that experience an increase in volatility.

Similarly, research by Bushee and Noe (2000) shows that institutional investors are more likely to invest in stocks with higher volatility, and an increase in their holdings often leads to an increase in volatility. Dennis and Strickland (2002) find that on days of large market returns, mutual funds evaluated on short performance can significantly impact stock prices more than individual investors. Factors such as growth investing and the rise of NASDAQ trading contribute to increased idiosyncratic volatility (Dennis & Strickland, 2002). Greenwood and Thesmar (2011) state that liquidity shocks associated with institutional ownership concentration or dispersion lead to increased volatility. Bushee (1998, 2001); Bushee and Noe (2000) utilise a classification for institutional investor types. They find that differences in investment strategies and preferences in dedicated institutions, quasi-indexers, and transient institutions impact firm management and future value distribution.

The literature on individual investor influence on volatility is extensive and focuses mainly on trading. Empirical studies by Kyröläinen (2008), Foucault et al. (2011), Aharon et al. (2022), and Baig et al. (2023) suggest that retail trading is associated with volatility. In contrast, Kostopoulos et al. (2022), Dorn and Huberman (2010), Baig et al. (2023) argue that ambiguity aversion, risk-aversion, and periods of crisis influence individual investor participation in the financial markets.

Kyröläinen (2008) found a positive relationship between individual investor trading and volatility. Their findings align with the theoretical noise trading models. Kyröläinen (2008) established that trading by individual investors is positively correlated with stock price volatility. This effect remains consistent across different periods and individual stocks. Findings suggest that less informed individual trades might impact volatility more than informed institutional trades.

In their study, Foucault et al. (2011) investigated how regulatory changes in France affected the market. The reforms raised trading expenses, resulting in a decrease in retail investor activity. As a result, volatility declined, suggesting that retail trading contributes positively to volatility. These results also hinted that the destabilising impact of retail momentum trades could outweigh the potentially stabilising effects of contrarian retail trades. Additionally, the study found that contrarian trades might contribute to volatility.

Along the same lines, Aharon et al. (2022) found that Robinhood's trading activity negatively affects the cross-listed securities' stability. Their findings are aligned with previous studies that demonstrated how uninformed investors can negatively impact the stability and efficiency of financial and capital markets. Aharon et al. (2022) found that increased Robinhood users' holdings in American depositary receipts are associated with increased volatility. A more recent study by Baig et al. (2023) found that individual investor marketable buy and sell retail trades lead to increased volatility during crises. Additionally, they discovered that these results are qualitatively similar when using an alternative proxy for retail trading, such as the growth in users on Robinhood.

Investor trading behaviour is influenced by ambiguity and risk aversion. Kostopoulos et al. (2022) examined the impact of ambiguity on individual investor trading. They discovered that there is more active trading and risk aversion during heightened ambiguity.

Their findings support the conclusions of Hirshleifer and Hong Teoh (2003) theory that investor sentiment becomes more important under ambiguous conditions. Thus, there is a connection between the level of individual investor ambiguity aversion and trading behaviour.

Lastly, Dorn and Huberman (2010) proposed the 'Preferred Risk Habitat Hypothesis,' which suggested that risk-averse investors might opt for lower-volatility stocks and lean towards mutual funds while heavily disregarding the return correlations in their portfolios. The results from their study, which used data from thousands of clients of a German brokerage, supported this hypothesis.

Both institutional and individual investors can impact volatility. However, there are differences in the ways they influence volatility. Institutional investors have a more dominant impact on days with large market returns, and factors such as growth investing contribute to increased idiosyncratic volatility. Individual investors contribute to volatility through their trading behaviours. More importantly, less informed individual trades impact volatility more than informed institutional trades.

2.5 Volatility measures

Volatility represents fluctuations in stock prices commonly driven by unforeseen news related to a company's performance (Venkatachalam, 2000). Volatility metrics derived from historical price data fall into three primary groups: absolute residuals, realised volatility, and historical volatility (Blasco, Corredor, & Ferreruela, 2012; Kyröläinen, 2008). As one of the components of mean-variance analysis, volatility is a focal point in financial analysis. Given the centrality of volatility to all claims within the hypotheses, understanding the differences in volatility measures and their quantification processes is important.

Realised volatility is vital in understanding market behaviours and how accurately volatility is estimated. Work by Merton (1980) set a precedent for accurately estimating volatility using fixed interval data, emphasising that shorter intervals reduce estimation errors, given that price movements follow a geometric Brownian motion. French, Schwert, and Stambaugh (1987) study highlighted the accuracy of monthly standard deviation estimators for stock market returns. Instead of rolling 12-month standard deviation estimators used in previous research, their methodology accounted for the non-constant nature of stock return volatility and autocorrelation in daily returns due to non-synchronous trading, offering a more precise estimation by focusing on within-month returns. Schwert (1989) extended this approach, using daily and weekly data to estimate monthly standard deviations in stock returns, reinforcing that non-overlapping samples minimise estimation errors through time. Andersen, Bollerslev, Diebold, and Ebens (2001) build on these earlier studies, emphasising the importance of high-frequency data in the construction of daily, rather than monthly, realised stock volatilities and underscoring the significance of advancing towards finer, high-frequency observations to estimate volatility more accurately. This method implies that as the time between observations shortens, the accuracy of estimation increases as continuous-time diffusion processes enable more precise measurements of volatility.

Various studies have delved into the concept of absolute volatility in stock returns. Schwert (1990) outlined the persistent nature of stock return volatility, its tendency to rise following market downturns, and its relationship with macroeconomic factors. Using a similar methodology to French et al. (1987), Schwert (1990) used past returns and day-of-the-week factors to estimate short-term changes in expected returns. They used absolute residuals to estimate stock return variability, assuming a normal distribution of returns. Following Schwert (1990), Jones, Kaul, and Lipson (1994), Chan and Fong (2000), and Avramov et al. (2006) each employed similar models that incorporated absolute residuals from regressions including lagged returns and day-of-the-week dummies to capture stock return volatility. These studies controlled for serial dependence in daily returns while exploring the effects of trade size, order imbalance, and sell-initiated transactions on daily stock volatility. Avramov et al. (2006) included normalised sell-initiated transactions in their regression analysis as a control variable. They aimed to understand the influence of trading dynamics on volatility.

Historical volatility measures, mainly through range-based measures, significantly evolved with contributions from Parkinson (1980) and Garman and Klass (1980). Parkinson (1980) introduced an approach for estimating historical volatility by considering the maximum and minimum daily prices of assets, recognising that extreme price data could offer valuable insights into the volatility's behaviour, particularly in predicting future volatility due to its mean-reverting nature. Garman and Klass (1980) suggested a method to estimate historical volatility using opening and closing prices and extreme prices. Their method reflects an understanding that detailed market movement data, like range-based and extreme data points, can estimate volatility more accurately than models using just open and close prices. Thus, capturing all intraday price movements is important for reliable volatility estimation. Subsequent studies and methodologies, such as those by Beckers (1983), Rogers and Satchell (1991), Andersen and Bollerslev (1998), Yang and Zhang (2000), and Alizadeh, Brandt, and Diebold (2002), further validated and expanded upon the initial concepts introduced by Parkinson (1980) and Garman and Klass (1980).

Kyröläinen (2008) argues that range-based volatility is a better proxy than absolute or squared returns. Range-based volatility measures have lower variance in measurement errors and accurately reflect intraday price variations. Range-based measures are more robust against bid-ask bounce. Thus, they are a more efficient volatility proxy than realised volatility, which can be biased due to the sum of squared high-frequency returns. Baig et al. (2023) found that the Alizadeh et al. (2002) range-based measure has a correlation of about 48% with the GARCH(1,1) volatility measure. Baig et al. (2023) propose that a range-based measure is a more dynamic measure of daily stochastic volatility. Hence, the range-based method is well suited for examining Robinhood users' behaviour due to its ability to account for intraday price variations, which matches with the pattern of active and daily monitoring of stocks commonly seen among Robinhood users.

3 Data and methodology

This chapter provides an overview of the data and methodology used in our study. We explain where the data came from and how we selected the sample for our analysis. Then, we explain the variables we derived from the data, including volatility, returns and individual investor ownership measures. Drawing on studies by Sias (1996), Aharon et al. (2022) and Welch (2022), our research approach combines non-parametric tests and regressions. We use methods such as the Kruskal-Wallis test, Wilcoxon rank-sum test and linear regressions to examine how individual investor ownership relates to volatility. Sias (1996) guides our choice of research design, and variables are influenced by the decisions taken by Aharon et al. (2022) and Welch (2022). In Chapter 1, based on the three research questions, we formulated three hypotheses that will allow us to test how individual investor behaviour influences stock price volatility. These three hypotheses are then tested using two non-parametric methods and four linear regressions.

For the first hypothesis, we conduct three analysis phases concerning the role of company size and individual investor ownership in shaping stock price volatility. Initially, the Kruskal-Wallis test assesses differences in median individual investor ownership and volatility across market capitalisation deciles. Subsequently, the Wilcoxon rank-sum test compares whether two samples likely originate from the same population, facilitating an understanding of whether increased volatility correlates with higher individual investor presence irrespective of capitalisation. Two regression analyses are employed to explore further the relationship between volatility, capitalisation, and individual investor holdings. Regression analysis permits the control for company size and past volatility, explaining the impact of volatility associated with higher levels of individual investor interest. The ensuing two regressions offer insights into temporal and directional considerations, enabling control for the impact of stock returns. The third regression examines the second hypothesis concerning the influence of changes in volatility on individual investor stock ownership. Finally, the last regression addresses the third hypothesis regarding the extent to which levels of individual investor ownership serve as an indicator of volatility.

3.1 Data description

To study the relationship between individual investor interest and volatility, we gathered daily historical price data, Robinhood popularity data, and market capitalisations for a universe of stocks comparable to ones used in Barber et al. (2022), and Welch (2022). The research data in this thesis is drawn from three sources: Robintrack, Yahoo Finance, and Financial Modelling Prep. The primary dataset for our analysis comes from the Robintrack website [1]. We obtained historical daily price data from Yahoo Finance and Financial Modeling Prep, which includes adjusted close, high, and low prices and market capitalisations. Subsequently, we merged this with the Robintrack dataset.

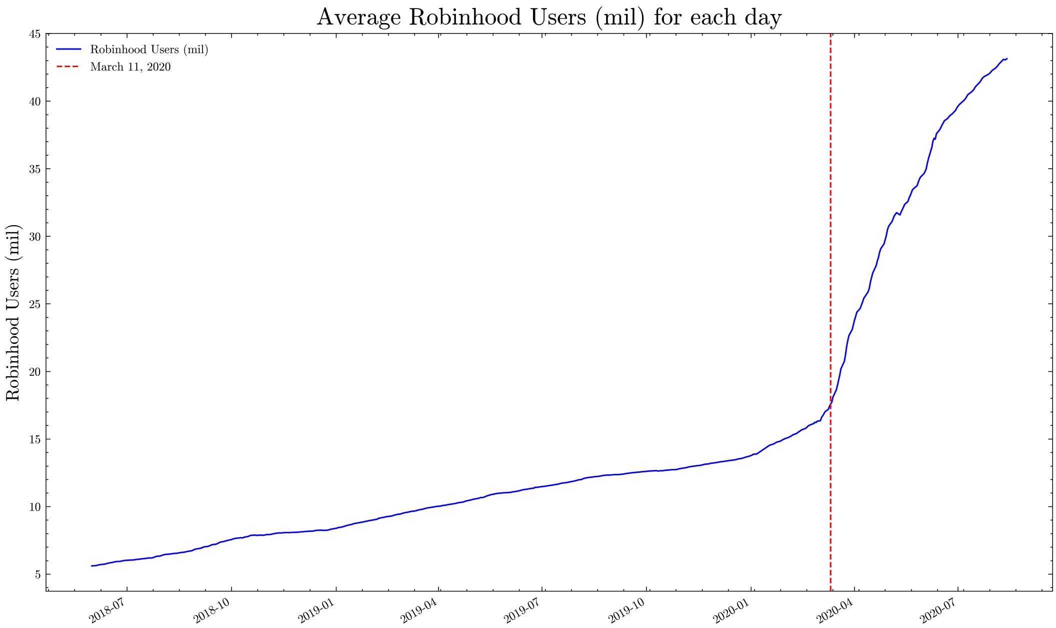

Figure 1. Robinhood total user stock holdings.

The rapid growth of Robinhood's user base is illustrated in Figure 1. The red dotted line indicates March 11 2020, a day the World Health Organisation assessed that the coronavirus outbreak had become a global pandemic (World Health Organization, 2020). Significant growth in popularity and usage of the platform can be observed to begin at Robinhood during this time.

[1] Robintrack data available for download at https://robintrack.net/data-download

Robintrack began collecting popularity metric data made publicly available on Robinhood's API on May 2, 2018. Robinhood shut down its public API on August 13, 2020. The popularity metric in the dataset signifies the count of unique accounts holding at least one share of a particular asset. However, it is worth noting that this data set only considers long shares and does not account for options.

Additionally, Figure 1 illustrates the potential endogeneity issue between user base growth and the number of holders for a particular stock. Especially when Robinhood marketing methods are considered, the growth in the user base may, partially or entirely, drive the increase in the number of holders for a specific stock.

3.2 Sample description

Barber et al. (2022) provide an overview of the Robintrack dataset. The Robintrack dataset contains 5,777,002 ticker-day observations for 802 unique days and 8,560 tickers. The mean stock has about 2,000 users, and the median user count for a stock is 160. The average day has 7,211 stock holdings, and just under 15 million user positions; user changes are generally small. Based on the number of users reported, Robinhood users have about three stock positions on average.

The sample utilised in our thesis represents a subset of the Robintrack dataset, explicitly focusing on ordinary equity stocks traded on Robinhood. Thus, this sample does not include exchange-traded funds, options, or cryptocurrencies available for clients to trade at Robinhood. After accounting for missing data, the sample used in the thesis contains 3765 tickers that can be classified as ordinary equity. Of the 802 unique days, 546 were trading days, and the final sample results were 1,721,810 ticker day observations.

3.3 Derived variables

The study's main variables include market capitalisation (Cap), volatility (σ), returns (r) and popularity weight (w). Except for market capitalisation, variables are derived from the data collected from the analysis period.

3.3.1 Volatility

Based on the advantages outlined by Kyröläinen (2008) and Baig et al. (2023), a log range volatility measure is chosen as it is more robust against measurement errors and accurately reflects intraday price variations. Following Alizadeh et al. (2002) and Gallant, Hsu, and Tauchen (1999), the log range volatility measure is created using daily historical price data, where volatility ($\sigma_{S,t}$) is the first logarithmic difference between the high and low price and H_t and L_t denote the high and low price at day $_t$.

$$\sigma_{S,t} = ln(H_t) - ln(L_t) \tag{1}$$

Baig et al. (2023) use the range-based method as a volatility proxy in their study on retail trading's impact on market stability. Similarly, Aharon et al. (2022) use a range-based volatility estimation to represent volatility to investigate Robinhood holdings' impact on American depositary receipts' volatility.

3.3.2 Returns

Returns calculated from adjusted close prices represent the percentage change in price from one trading day to the next, considering adjustments for stock splits and dividend distributions. Following Center for Research in Security Prices standards, the adjusted close price accounts for stock splits and dividend adjustments, ensuring accurate historical data analysis and consistency and reliability in calculations.

$$r_t = \frac{AdjC_t}{AdjC_{t-1}} - 1 \tag{2}$$

Thus, the formula calculates the daily simple return (r_t) by dividing the current day's adjusted closing price ($AdjC_t$) by the previous day's adjusted closing price ($AdjC_{t-1}$) and subtracting one from the result to determine the return percentage. While log-range volatility focuses on the price fluctuations seen during intraday trading sessions, returns calculated from adjusted close prices also reflect overnight returns. As a result, volatility and returns can exhibit different magnitudes, revealing different aspects of price-related market dynamics.

3.3.3 Popularity weight

The study by Sias (1996) focused on institutional investors. They measured institutional investor participation by examining a fraction of the shares they held each September. This approach also indirectly reflects the portion of shares owned by individual investors. The Robintrack dataset does not allow us to replicate this measure directly. However, prior research offers a suitable solution that addresses this limitation.

We opted for a proxy of individual investor ownership. The chosen measure is similar to one Welch (2022) used, where individual stock holds weight in the 'ARH portfolio'. Thus, the measure is comparable to the percentage holding in an aggregate individual investor portfolio and results in an adequate proxy for a fraction of shares, which is the method Sias (1996) utilised. Thus, the portfolio is not a market capitalisation-weighted portfolio but reflects the active choices of individual investors. Furthermore, Welch (2022) states that this representation is a 'very good proxy' for a household equal-weighted portfolio.

To measure individual investor ownership, we create a variable called 'popularity weight' using Robintrack popularity data. In popularity weight, the count of unique accounts holding a minimum of one share of a particular asset is weighted to the total number of accounts holding a minimum of one share of any asset on Robinhood. Thus, popularity weight, represented by the variable w, is the weight $w_{i,t}$ in the portfolio, i is a given stock, and N_i is the number of Robinhood holders in stock.

$$w_{i,t} = \frac{N_{i,t}}{\sum_i N_{i,t}} \tag{3}$$

Alternatively, the method by Aharon et al. (2022) could have been chosen to capture the impact of Robinhood holdings. Aharon et al. (2022) utilised two variables: level measure and growth measure. The level measure captures the daily average number of intraday Robinhood accounts holdings. The growth measure captures the growth rate of the daily average number of intraday Robinhood account holdings. These measures are then used to capture how trader activity impacts volatility.

The main disadvantage of the Aharon et al. (2022) method is that the level measure or the growth rate measure does not consider the growth rate in the Robinhood user base. Another limitation of this approach would have been using two separate variables to capture individual investor ownership. Therefore, we chose an alternative method to the level and growth-based variables. Weight-based measure corrects for the issue of growing user base by design and popularity weight variable is more adept at capturing individual investor ownership. Moreover, according to Welch (2022), this representation can be viewed as a ‘consensus statistic’, a ‘crowd wisdom’ measure or a monotonic transformation of ‘breadth of ownership’.

3.4 Data exploration

Before establishing regression models, we explore whether firm-specific stock volatility can predict subsequent changes in individual investor ownership. This assessment serves two primary purposes. First, check if the evidence matches overall ownership patterns and how investors behave. Second, we need to understand the temporality of the relationship between volatility and Robinhood holdings to determine how long the lags should be in our regression models. Following Welch (2022) methodology, we present two tables with panels A and B. Panel A shows Robinhood holders and changes in holdings for each date and ticker. Panel B illustrates changes in Popularity Weight, Rate of Return, and Volatility for each company over three days. The differences between thesis tables and tables in Welch (2022) study can be explained by differences between samples and the method of selecting data for the tables: using functions from the pandas library by Wes McKinney (2010) in thesis and hand-picked in Welch (2022).

The 'Extreme one-day increases in Robinhood holdings' table highlights stocks that experienced the most significant one-day increases in popularity weight. On the other hand, the 'Extreme one-day decreases in Robinhood holdings' table presents stocks with the most significant one-day decreases in popularity weight. We expect volatility to be a more recurring characteristic than substantial one-day increases or decreases in the stock price as suggested by Welch (2022). The tables show results that are mainly in line with our expectations.

Table 1 shows ten stocks that experienced unusually large one-day increases in popularity weight. The recurring characteristic seems to be above median volatility for the respective market capitalisation decile of the company. Table 1 suggests a contrarian pattern in trading behaviour, where extreme increases in popularity weight follow significant negative returns that coincide with high volatility. Market capitalisations of companies could potentially influence the level of these increases.

Table 2 lists ten stocks that experienced unusually large one-day decreases in popularity weight. It can be observed that large decreases often coincide with increased volatility that persists for three days. Table 2 suggests a pattern of individual investors' behaviour corresponds with the disposition effect, in which Robinhood users sell or continue to sell stocks after high one-day stock price increases.

Table 1. Extreme one-cay increases in Robinhood holdings.

Panel A shows Robinhood holders in thousands for each date and ticker and the change in holdings ($\Delta{-1,0}$). Panel B shows a change in Popularity Weight, Rate of Return, and Volatility for each company, including the initial popularity weight (w_0), change in popularity weight ($\Delta{-1,0w}$), rate of return for three days (r_{-1}, r_0, r_{+1}), and volatility for the same period ($\sigma{-1}$, σ_0, σ_{+1}).

Panel A: Robinhood Holdings in Thousands

	Date	Ticker	-2	-1	±0	$+1$	$+2$	$\Delta_{-1,0}$
1	July 26, 2018	FB	108	114	156	166	170	42,083
2	April 4, 2020	UBER	144	144	196	196	196	52,329
3	July 29, 2020	KODK	9	34	119	122	105	84,394
4	January 6, 2020	INPX	0	0	26	26	26	26,032
5	October 2, 2018	OGEN	17	27	41	39	37	14,336
6	July 17, 2018	NFLX	101	107	118	118	117	11,000
7	September 19, 2019	TLRY	18	18	31	31	31	12,368
8	June 29, 2020	VXRT	22	22	81	86	55	58,475
9	July 15, 2020	MRNA	214	223	284	305	315	60,914
10	July 27, 2018	FB	114	155	166	170	173	9,832

Panel B: Robinhood Change, Rate of Return and Volatility

		Popularity Weight		Rate of Return			Volatility		
	Company name	w_0	$\Delta_{-1,0}w$	r_{-1}	r_0	r_{+1}	σ_{-1}	σ_0	σ_{+1}
1	Facebook	2.49	0.65	-0.7	−18.9	1.3	3.9	3.60	2.0
2	Uber	0.82	0.21	-6.8	8.1	0.3	9.1	7.59	4.8
3	Kodak	0.28	0.20	-10.1	123.2	203.0	47.9	123.3	43.7
4	Inpixon	0.18	0.18	8.88	-5.6	38.29	16.5	122.0	49.5
5	Oragenics	0.55	0.18	-50.9	14.6	39.1	54.9	64.1	41.3
6	Netflix	0.19	0.17	-1.1	45.5	1.1	28.5	11.2	2.9
7	Tilray	0.42	0.16	-17.9	95.9	28.9	43.4	68.3	16.1
8	Vertex Pharmaceuticals	0.20	0.14	-18.1	-0.1	28.4	43.2	30.0	59.4
9	Moderna	0.69	0.14	1.9	9.6	4.5	5.7	12.4	8.5
10	Facebook	2.63	0.14	-2.1	-0.7	-18.9	5.11	3.9	3.6

Table 2. Extreme one-cay decreases in Robinhood holdings.

Panel A shows Robinhood holders in thousands for each date and ticker and the change in holdings ($\Delta{-1,0}$). Panel B shows a change in Popularity Weight, Rate of Return, and Volatility for each company, including the initial popularity weight (w_0), change in popularity weight ($\Delta{-1,0w}$), rate of return for three days (r_{-1}, r_0, r_{+1}), and volatility for the same period ($\sigma{-1}$, σ_0, σ_{+1}).

Panel A: Robinhood Holdings in Thousands

	Date	Ticker	-2	-1	± 0	$+1$	$+2$	$\Delta_{-1,0}$
1	February 27, 2019	IGC	78	78	42	42	42	-36,098
2	November 2, 2018	INPX	39	40	1	1	1	-35,583
3	February 5, 2020	TSLA	154	163	148	152	152	-15,439
4	October 1, 2018	TSLA	95	101	94	94	94	-6,876
5	October 4, 2018	IGC	89	84	76	72	80	-7,259
6	June 10, 2020	TSLA	85	85	80	73	76	-4,857
7	January 16, 2020	FIT	257	256	256	257	259	-924
8	October 24, 2018	TSLA	113	109	103	98	95	-5,885
9	October 03, 2018	IGC	81	89	83	76	72	-5,551
10	April 8, 2019	STZ	7	7	0	7	7	-7,584

Panel B: Robinhood Change, Rate of Return and Volatility

		Popularity Weight		Rate of Return			Volatility		
	Company name	w_0	$\Delta_{-1,0}w$	r_{-1}	r_0	r_{+1}	σ_{-1}	σ_0	σ_{+1}
1	India Globalization	0.44	-0.54	-3.8	1.8	13.4	11.4	20.7	23.2
2	Inpixon	0.02	-0.48	-2.4	-32.1	2.3	31.5	69.4	14.8
3	Tesla	0.97	-0.10	13.7	-17.1	1.9	15.0	18.3	14.7
4	Tesla	1.25	-0.10	-13.9	-17.3	-3.1	6.4	3.3	5.7
5	India Globalization	1.00	-0.09	-30.1	27.5	-36.8	23.1	21.4	29.6
6	Tesla	1.24	-0.07	0.9	16.1	-0.3	3.3	7.9	3.5
7	Fitbit	1.76	-0.07	0.9	0.3	1.3	1.7	1.5	1.5
8	Tesla	1.31	-0.07	12.7	-0.1	9.1	12.8	6.3	6.4
9	India Globalization	1.09	-0.07	44.1	-31.9	-27.5	42.8	23.1	21.4
10	Constellation Brands	0.00	-0.07	0.2	0.5	0	2.37	1.0	1.4

3.5 Methodology

The following section will detail the reasons behind selecting particular methodologies informed by existing research studies. We will review each method's thesis goals and aims and compare the research objectives to established theories, concepts, and findings mentioned in the literature review. Additionally, we will outline the expected results for each method, in line with the hypotheses from Chapter 1.

Our methodology involves two main steps. First, we will use non-parametric analysis techniques to study the connections between market capitalisation, volatility, and popularity weight. Second, we will conclude with regression analysis. The two parametric methods are the primary analytical techniques used in our study. At the same time, the regressions offer us information relating to temporal and directional concerns and additionally allow us to control for variables such as market capitalisation.

The analysis will be done using Python. Pandas library by Wes McKinney (2010) is used for data structures. Non-parametric methods use the SciPy library by Virtanen et al. (2020). Regressions are conducted using the statsmodels library by Seabold and Perktold (2010).

3.5.1 Kruskal-Wallis test

The statistical analysis of the aggregate begins with a non-parametric procedure that explores volatility and popularity weight based on market capitalisation deciles. Although this analysis follows a less frequently utilised methodology in financial literature, the test is one the most widely used non-parametric procedures for testing whether samples are from identical samples according to W. W. Daniel (1978).

A study by Sias (1996) used the Kruskal-Wallis test to explore the connection between institutional investor ownership and volatility. This thesis uses the test to check for significant differences in the median popularity weight and volatility across market capitalisation deciles. We test the first hypothesis about the impact of company size, volatility and individual investor ownership. If the test rejects the null hypothesis, there are significant variations in individual investor ownership and the volatility of stocks across capitalisation deciles.

We expect significant differences in the popularity weight and how volatile stocks are across market capitalisation deciles. Firstly, we expect the test to reveal an inverse relationship between market capitalisation and volatility. This expectation aligns with previous studies in finance, such as Sias (1996), and the size effect documented by Fama and French (1993), Liew and Vassalou (2000), and Campbell and Vuolteenaho (2004). These studies demonstrate that more significant, established companies exhibit lower volatility compared to smaller ones and vice versa. This inverse relationship suggests that factors such as increased liquidity and stability of cash flows contribute to lower volatility in larger companies.

Secondly, we anticipate that the volatility preferences of Robinhood investors will display an asymmetric pattern expressed among the market capitalisation deciles. These expectations are based on the literature on the propensity to gamble, risk and ambiguity aversion and individual investor diversification. While high-volatility investment options are prevalent on trading platforms, as Kumar (2009) suggested, attracting investors seeking sensations akin to gambling, Kostopoulos et al. (2022) find investors tend to exhibit less risk-taking behaviour during periods of increased ambiguity. Additionally, Dorn and Huberman (2005, 2010) show that many investors maintain minimally diversified portfolios and systematically select lower-volatility stocks, suggesting a preference for stability.

3.5.2 Wilcoxon rank-sum test

A second non-parametric test follows the initial non-parametric test, the Wilcoxon rank-sum test. While the Kruskal-Wallis tests examine variations in individual investor holdings and volatility across various company sizes, we cannot conclude whether increased volatility correlates with a greater individual investor presence regardless of market capitalisation. This test is an additional test for the first hypothesis.

Following the methodology by Sias (1996), the data is divided into two categories within each capitalisation decile based on popularity weight. Stocks with weights greater than the overall median weight are classified as high weight, while those with weights less than the median are classified as low weight, with the median omitted from each decile. Wilcoxon rank-sum test is then used to compare whether two samples are likely to come from the same population (W. W. Daniel, 1978). The Wilcoxon rank-sum test compares two groups of stocks: one with a low popularity weight and another with a high popularity weight within each decile. It assesses whether there is a significant difference between these groups.

We anticipate a significant difference between these groups for volatility, which implies notable differences in volatility between the low and high-popularity weight groups within each decile. Additionally, we expect higher volatilities in all high-popularity weight group deciles, indicating that individual investors are attracted to stocks with higher volatility. A possible reason is the perceived higher potential returns. The second expectation is based on findings by Kumar (2009), which suggests that individual investors are attracted to volatility. Additionally, based on the findings by Sias (1996), we do not expect differences in preferences for volatility between individual and institutional investors, allowing for changes in the degree of preference.

3.5.3 Linear regressions

In the two non-parametric analyses, we were concerned whether the medians of popularity weight and volatility across deciles are significantly different. We now move to analyse the mean, specifically the conditional mean. Linear regression estimates the conditional mean of the dependent variable. Ordinary least squares is the most common technique for estimating coefficients of linear equations. The regression line is estimated to minimise the sum of the squared differences between the observed values and the values predicted by the line (Wooldridge, 2013).

In related studies, both Sias (1996) and Aharon et al. (2022) used linear regressions. The study by Sias (1996) included three regressions with a maximum of three regressors. Regressors were a mixture of volatility, ownership, returns and company size variables. Aharon et al. (2022) used OLS regression on a pooled sample of American depositary receipt day observations with robust standard errors. Aharon et al. (2022) regressions were done without fixed effects and with models with various combinations of fixed effects for country, time and industry.

OLS regression assumes a transparent population model with parameters to estimate, random sampling, no perfect collinearity among explanatory variables, mean independence of the error term, zero population mean for the error, consistent error variance across all data points and normality of the error term (Wooldridge, 2013). However, collinearity and autocorrelation are common issues with financial markets or time-series datasets. In the thesis' data, for instance, stock returns may closely correlate with volatility, and larger companies with higher market capitalisations may have higher popularity weights.

The problem of collinearity implies that there are redundant independent variables, meaning that the same information is provided in another way. Variance inflation factors (VIF) is a standard measure of collinearity that offers a simple diagnostics for detecting overall collinearity problems that do not involve the intercept. VIF tells us that the higher the proportion of variance that is not explained by the other independent variables. Different guidelines exist for interpreting VIF values, with some suggesting a threshold of 10. Autocorrelation in the residuals will be detected using the Durbin–Watson test for independence. The test is for first-order autocorrelation, which tells us whether residuals are related to their immediately preceding values. It is a ratio of the squared differences between adjacent residuals to the squared sum of all residuals. (Brooks, 2014; Rawlings, Pantula, & Dickey, 1998)

To continue the analysis that began with the Kruskal-Wallis test and the Wilcoxon rank-sum test about the relationship between volatility, capitalisation, and individual investor holdings, we will regress volatility on the natural logarithm of capitalisation and the popularity weight. The regression will examine whether a company's market capitalisation and the level of individual investor interest significantly impact its volatility. The first two regressions test the first hypothesis of whether company size and the level of individual investor ownership significantly impact volatility.

The first regression models the relationship between volatility (σ), market capitalisation (Cap) and popularity weight (w). It suggests that a stock's volatility is a function of the log of the market capitalisation and popularity weight.

$$\sigma_t = \alpha + \beta_1 ln(Cap_t) + \beta_2(w_t) + \varepsilon_t \tag{4}$$

The second regression, in addition to market capitalisation (Cap) and popularity weight (w), incorporates lagged values of volatility from one to three days, denoted as σ_{t-1}, σ_{t-2}, and σ_{t-3}. The equation for the second regression is as follows:

$$\sigma_t = \alpha + \beta_1 ln(Cap_t) + \beta_2(w_t) + \beta_3(\sigma_{t-1}) + \beta_4(\sigma_{t-2}) + \beta_5(\sigma_{t-3}) + \varepsilon_t \tag{5}$$

With the first two regressions, We want to determine whether higher levels of individual investor interest are associated with higher volatility after controlling for market capitalisation. The results are expected to show that, after controlling for market capitalisation, higher levels of volatility are associated with higher levels of individual investor ownership. This association would align with the claim in the first hypothesis. The results would also indicate that securities with greater volatilities may attract individual investors, and an increase in individual investor interest induces an increase in volatility. Thus, the results are expected to align with findings from Aharon et al. (2022) study conducted with a smaller sample limited to American depositary receipts.

We test the second hypothesis after exploring the contemporaneous relationship between Robinhood ownership and market capitalisation. We examine the relationship between change in volatility and individual investor ownership. The third regression will analyse whether a stock's past performance and perceived risk influence the current level of individual investor ownership. It models the relationship between the current popularity weight (w) and changes in volatility ($\Delta\sigma$), popularity weight ($w_t - 1$), and returns ($r_t - 1$). The regression suggests that the current stock's popularity weight is a function of change in volatility, past popularity weight, and past returns.

$$w_t = \alpha + \beta_1 \Delta\sigma + \beta_2(w_{t-1}) + \beta_3(r_{t-1}) + \varepsilon_t \tag{6}$$

We want to find out whether changes in volatility can explain popularity weight in the Robinhood portfolio beyond what is explained by lagged popularity weight and returns. That is, we hope to find that results provide evidence that an increase in volatility attracts individual investors. Thus, the expected results are aligned with the conclusions based on Kumar (2009). Additionally, the regression examines the relationship between popularity weight and returns. We expect this relationship to be inverse, which would align with the literature on contrarian trading behaviour. Thus, the expectations for parametric tests are not different from those for non-parametric tests.

The final regression concerns the changes in Robinhood holdings and volatility and tests the third hypothesis. Additionally, the regression is conducted to address the issue of reverse causality, which concerns whether increased interest from individual investors results in heightened volatility or vice versa. To test whether popularity weight is associated with a subsequent change in volatility, we regressed the volatility in day t on the volatility of day t_{-1}, the daily lagged change in popularity weight and the return in day t_{-1}. That is, we examine whether lagged change popularity weight can explain variation in volatility beyond that explained by lag volatility. We included models with two- and three-day lags based on the expected results of the extreme changes in Robinhood holdings tables. Results for additional models, model (2) with a lag of 2 (t_{-2}) and lag of three days (t_{-3}) for change in popularity weight, are presented together with the base model:

$$\sigma_t = \alpha + \beta_1 \sigma_{t-1} + \beta_2(\Delta w) + \beta_3(r_{t-1}) + \varepsilon_t \tag{7}$$

Thus, the regression models the relationship between current volatility (σ) and past volatility σ_{t-1}), change in popularity weight(Δw), and past returns (r_{t-1}). The model suggests that a stock's current volatility is a function of past volatility, change in popularity weight, and past returns.

The regression will examine whether popularity weight is associated with subsequent changes in volatility. We hope to find out whether lagged changes in popularity weight can explain variation in volatility beyond what is explained by lagged volatility alone. Therefore, the results will indicate whether an increase in individual interest induces an increase in volatility.

3.6 Limitations

This thesis has limitations related to the representativeness of the sample related to individual investors and the comparability of the results related to institutional investors.

Robintrack dataset represents data about only one brokerage's users. As stated by Aharon et al. (2022), not every individual investor is a Robinhood client, and not every Robinhood client is necessarily an individual investor. However, as pointed out by Welch (2022), as the number of Robinhood users increased, so did the reliability of the statistics of Robinhood investors holding stocks. Robinhood measured user volume; the brokerage was a significant competitor and accounted for 30% of the retail trading volume during the sample period. Thus, given the law of large numbers, the Robintrack dataset should represent individual investors.

The thesis has a broader stock universe and a shorter measuring interval than used in the study by Sias (1996). Sias (1996) gathered weekly returns, annual institutional holdings, and annual market capitalisations for securities listed on the New York Stock Exchange. This thesis sample contains daily data on individual investor holdings and covers ordinary stocks from a broader set of exchanges: NASDAQ, NYSE, and NYSE American.

Concerning the comparability between the thesis' results and Sias (1996) findings for institutional investors, there are two potential issues related to the time frame and volatility measures. This thesis uses daily data, while comparable data on institutional investor ownership is an annual snapshot of a specific day for each year. Although both studies explore volatility, the measures utilised differ, as Sias (1996) uses realised volatility. Arguably, range-based volatility is a more efficient volatility proxy for an intraday sample than realised volatility Baig et al. (2023); Kyröläinen (2008). Therefore, due to these limitations, the comparisons should not be made regarding the magnitude or scale of volatility.

4 Empirical results and discussion

In this chapter, we present empirical results and discuss the findings. We start with descriptive statistics. After, we move to the results from the analysis conducted based on the methodology set in Chapter 3. We begin with results from the two non-parametric procedures. We examine Kruskal-Wallis test results and follow with Wilcoxon rank-sum test results. Then, we present the regression results. First, we explore the results of two regressions concerning Robinhood ownership and company size. The two non-parametric procedures and two regressions answer the first hypothesis about the interaction between company size, individual investor ownership and volatility. Second, we present regression results concerning changes in volatility and Robinhood holdings. These results will answer the second hypothesis about the relationship between change in volatility and individual investor ownership. Then, we move on to the final regression result, which concerns the predictability of individual investor ownership as an indicator of volatility. After the empirical results, we discuss them in greater detail and relate the resutls to the empirical findings and theory presented in the literature review.

4.1 Descriptive statistics

Descriptive data statistics were generated for all variables. Descriptive statistics for the sample are highlighted in Table 3, which include the mean, median, standard deviation, and the 25th and 75th percentiles. The statistics highlighted in the table mirror those that Aharon et al. (2022) presented, allowing for comparisons.

The data shown in Table 3 indicates that the average and middle values for popularity weight, volatility and changes in popularity weight and volatility are similar. Similarity hints at normal distribution. Moreover, the standard deviations for these measures imply a relatively consistent range of values around their means and medians. In this dataset, the 25th and 75th percentiles for popularity weight, volatility and changes in popularity weight and volatility suggest a consistent spread around the middle values.

Table 3. Descriptive statistics.

	Popularity weight	Volatility	Return	Δ Popularity weight	Δ Volatility
Mean	0.00021	0.047	0.00058	-7.15E-08	2.19E-05
Median	2.42E-05	0.034	0	-4.31E-08	-0.00015
std	0.0012	0.048	0.051	1.59E-05	0.041
25%	7.01E-06	0.020	-0.015	-3.19E-07	-0.011
75%	8.10E-05	0.058	0.014	4.90E-08	0.010

The uniformity, coupled with the close alignment of mean and median values, implies a symmetric distribution with moderate variability. However, It's important to note that the data in the sample doesn't follow a perfect normal distribution. Even though the mean and mean values are close, there are deviations from complete normality, as shown by minor differences. Moreover, skewness, especially in how popularity weight and volatility are distributed, indicates that the data might show some non-normal characteristics.

Comparing the volatility statistics presented in the thesis's descriptive statistics with those documented in the study by Aharon et al. (2022) reveals both similarities and differences. Both volatility trends have similar mean and median, with Aharon et al. (2022) reporting slightly higher mean volatility. There are also slight variations in volatility as the thesis has a narrower dispersion than Aharon et al. (2022). Overall, volatilities between the two samples seem to align closely.

Table 4 illustrates the correlations between key variables. It is interesting to note that returns are positively correlated with volatility and change in volatility, albeit moderately. Meanwhile, volatility exhibits a stronger positive correlation with change in volatility, indicating a potential pattern of volatility persistence. On the other hand, popularity weight shows weak correlations with other variables, suggesting its relative independence.

Table 4. Pairwise correlations of variables.

	Returns	Popularity weight	Volatility	Δ Popularity weight	Δ Volatility
Return	1	0.0028	0.14	0.089	0.144
Popularity weight		1	-0.0044	-0.037	-0.00063
Volatility			1	0.094	0.43
Δ Popularity weight				1	0.053
Δ Volatility					1

While some correlations exist, they remain modest, indicating a lack of multicollinearity among the variables. However, multicollinearity is better assessed through variance inflation factors in a regression context. For variance inflation factors, the dataset did not have problems with multicollinearity—the highest VIF 2, which indicates no significant collinearity in the regressions.

However, autocorrelation poses issues. Durbin-Watson's statistic of 0.867 indicated potential autocorrelation in regression Model 1 concerning volatility, market capitalisation, and popularity weight. However, the regression Model 2 with additional lagged volatility had no autocorrelation, as the test statistic value 2 indicated. Regression Model 3, examining the relationship between changes in volatility and Robinhood holdings, had a test statistic value of 1.378, suggesting a degree of positive autocorrelation. However, this is close to the acceptable range of 1.5 to 2.5 for large samples with a moderate number of regressors (Turner, 2020).

The Newey-West estimator can address autocorrelation and produce more accurate standard errors and coefficient estimates. Newey and West (1987) created a variance-covariance estimator that stays accurate even when heteroscedasticity and autocorrelation (Brooks, 2014). Heteroskedasticity-consistent standard errors by White (1980) could be used as they were in the study by Sias (1996). However, White's correction fixes for heteroscedasticity, and the Newey-West estimator corrects for both autocorrelation and heteroscedasticity by giving heteroscedasticity and autocorrelation consistent standard errors.

4.2 Non-parametric procedures

The statistical analysis of the aggregate begins with two non-parametric procedures. They investigate the relationship between volatility and popularity weight by categorising them into deciles based on market capitalisation and popularity weight sorted portfolios. These two tests, in conjunction, test the impact of market capitalisation and the level of individual investor interest on volatility.

4.2.1 Kruskal-Wallis test results

The results of the Kruskal-Wallis test are broadly consistent with our expectations. Table 5 presents median popularity weight and volatility categorised by market capitalisation deciles, ranging from the smallest to the largest. The chi-square test statistic indicates statistically significant differences among the median popularity weight and volatility across capitalisation deciles.

Table 5. Median popularity weight and volatility by capitalisation.

Market Capitalisation Decile	Median Popularity Weight	Median Volatility
Smallest	0.0041%	6.76%
Decile 2	0.0019%	5.31%
Decile 3	0.0012%	4.60%
Decile 4	0.0016%	4.10%
Decile 5	0.0018%	3.73%
Decile 6	0.0016%	3.35%
Decile 7	0.0018%	3.02%
Decile 8	0.0021%	2.65%
Decile 9	0.0037%	2.36%
Largest	0.0152%	1.96%
The chi-square Statistic (all equal)	242466.50***	319851.19***

***indicates statistical significance at the 1 per cent level.

The results of the Kruskal-Wallis test show that we can confidently reject the null hypothesis that there are no differences in popularity weight or volatility among stocks in each capitalisation decile with 99% confidence. Table 5 indicates a positive relationship between investor holdings and capitalisation and a negative relationship between volatility and capitalisation. Results show an inverse relationship between market capitalisation and volatility. Median volatility decreases from 6.76% for the smallest decile to 1.96% for the largest decile, aligning with previous studies in finance that describe a relationship between market capitalisation and volatility.

As noted previously, however, the relationship between market capitalisation and individual investor interest may be motivated by reasons other than risk avoidance. The median popularity weight shows an intriguing pattern contrasting individual and institutional investors as Sias (1996) reported. While popularity weight tends to increase with market capitalisation from deciles three to the largest, individual investors favour two groups of stocks in the smallest deciles, showing high median volatility and popularity weight. This deviation is consistent with conclusions drawn from studies by Kumar (2009) and Dorn and Huberman (2010). We see a preference for both high-volatility and low-volatility stocks.

4.2.2 Wilcoxon rank-sum test results

The Kruskal-Wallis test results suggest that a stock's capitalisation has a statistically significant impact on its popularity, weight, and volatility. We sorted each observation into high or low-popularity weight groups to study the relationship between volatility and individual investor interest while keeping capitalisation constant.

The Wilcoxon rank-sum test results for popularity weight-sorted portfolios have confirmed our expectations, and the results align with conclusions made based on the literature review. Significant differences were observed between groups based on popularity weight regarding both capitalisation and volatility. Noticeable variations were evident in company size and volatility between the low and high-popularity weight groups within each decile. Additionally, higher volatilities were observed in the high popularity weight groups, supporting the conclusion that individual investors are drawn to stocks with greater volatility. This preference for higher volatility is potentially due to perceived higher potential returns or other factors.

Table 6. Popularity weight sorted portfolios.

This table presents the distribution of securities across high and low-popularity weight categories within each capitalisation decile. This table sorts stocks into high-weight ("High") or low-weight ("Low") categories based on their popularity relative to the overall median weight. The median is excluded from each decile for classification.

	Popularity Weight %			Market Capitalisation			Volatility %		
Deciles	High	Low	Z-stat	High	Low	Z-stat	High	Low	Z-stat
Smallest	0.01	0.001	355.9***	20163	24139	-48.5***	7.65	5.81	81.8***
Decile 2	0.006	0.0006	355.9***	78595	79078	-7.1***	6.22	4.30	104.9***
Decile 3	0.005	0.0003	355.9***	184642	185036	-5.6***	5.74	3.41	148.0***
Decile 4	0.006	0.0004	355.9***	345222	336151	19.4***	5.04	3.15	137.2***
Decile 5	0.006	0.0005	355.9***	621669	628250	-12.6***	4.47	3.05	118.3***
Decile 6	0.005	0.0006	355.9***	1141989	1126469	16.7***	4.05	2.70	125.6***
Decile 7	0.006	0.0007	355.9***	1987486	1969895	6.3***	3.52	2.57	102.0***
Decile 8	0.006	0.0008	355.9***	3694125	3565540	23.6***	3.08	2.28	98.6***
Decile 9	0.011	0.001	355.9***	8682011	7632167	54.4***	2.70	2.08	87.9***
Largest	0.058	0.005	355.9***	54294630	24367425	176.4***	2.03	1.90	23.3***

*** indicates statistical significance at the 1 per cent level.

Table 6 presents a comprehensive analysis of the distribution of securities across high and low popularity weight categories within each capitalisation decile, alongside market capitalisation and volatility metrics. These tests examine the hypothesis that, within each capitalisation decile, there is no significant difference between the low- and high-popularity weight classifications regarding individual holdings, capitalisation, or volatility. The Wilcoxon rank-sum test results show statistically significant differences in popularity weight for stocks in the low- and high-popularity weight categories. Based on the z-statistics from the Wilcoxon rank-sum tests with 99% confidence, we can reject the null hypothesis that there is no significant difference between the two popularity weight classifications.

A comparison of volatilities for stocks in each of the high- and low-weight groups shows that high-weight groups have higher volatility across all deciles. Thus, individual investor ownership measured by popularity weight is associated with increased stock volatility, regardless of market capitalisation. Individual investor ownership shows similarities to institutional investor ownership. The findings align with previous volatility research by Sias (1996) regarding institutional investors, which indicates that after controlling for capitalisation, higher volatility is correlated with the presence of institutional investors.

Further examination of market capitalisations supports the asymmetric risk preferences observed in Table 5. Interestingly, while the highest popularity weights are primarily found in the largest two market capitalisation deciles, the smallest two deciles, both the lowest and highest classifications, have high popularity weights. Additionally, Robinhood investors prefer smaller market capitalisation stocks in smaller market capitalisation deciles and larger market capitalisation companies in larger capitalisation deciles.

The findings from the Kruskal-Wallis and Wilcoxon rank-sum tests can be attributed to two distinct behaviours observed among individual investors in the literature review. Firstly, Robinhood investors could prefer stocks that resemble lotteries based on the characteristics outlined by Kumar (2009). Weiss-Cohen et al. (2024) identify two behavioural patterns among gamblers, both present among Robinhood investors, who prefer low- and high-volatility stocks. Secondly, the preference for low-volatility stocks among Robinhood investors may be explained by ambiguity aversion, as suggested by Kostopoulos et al. (2022) and the preference for low volatility of risk-averse investors, as highlighted by Dorn and Huberman (2010).

On average, Robinhood investors hold three stocks in their portfolios, indicating a tendency towards under-diversification. Dorn and Huberman (2005) and Mitton and Vorkink (2007) highlight the prevalence of under-diversified portfolios among investors due to challenges in optimisation implementation and a propensity towards ’narrow framing’. Furthermore, some investors navigate the trade-off between skewness and diversification by opting for under-diversified portfolios to pursue high volatility and potentially higher returns.

4.3 Regression analysis

Contemporaneous regression analysis was done in three phases. First, we examined Robinhood ownership and market capitalisation about volatility. Second, we analysed changes in volatility and Robinhood holdings. Third, we studied the impact of the changes in Robinhood holdings on volatility.

While the two parametric methods allowed us to study whether individual investors are attracted to volatility, the four following regressions allowed us to analyse information relating to temporal and directional concerns and to control for important variables such as market capitalisation and returns. All the regressions use heteroskedasticity and autocorrelation-consistent standard errors.

4.3.1 Robinhood ownership and company size

We continue the analysis presented in Table 5 and Table 6 with the results of a contemporaneous regression of volatility on the natural logarithm of capitalisation and the popularity weight of the Robinhood portfolio.

The results presented in Table 7 suggest that higher levels of volatility are associated with higher levels of individual investor interest after controlling for capitalisation. Specifically, the coefficient associated with Robinhood popularity weight is positive and statistically significant at the 1 per cent level.

Table 7. Contemporaneous regression analysis.

Variable	Coefficients	Std. Error	z	P-value
Intercept	0.1513***	0.000	415.875	0.000
$\ln$(Market Capitalisation)	-0.0077***	2.48e-05	-310.930	0.000
Weight	2.6718***	0.037	72.510	0.000
No. Observations	1721810			
R-squared	0.119			
Adj. R-squared	0.119			

*** denotes significance at the 1 per cent level.

In the additional model with lagged variables volatility, statistically significant relationships at the p = 0.01 levels are observed. Market capitalisation is negatively associated with volatility, suggesting that larger companies exhibit lower volatility than smaller companies. Conversely, popularity weight shows a positive relationship, indicating that higher levels of individual investor interest correspond to increased volatility.

Table 8. Contemporaneous regression analysis with lagged volatility.

Variable	Coefficients	Std. Error	z	P-value
Intercept	0.0412***	0.000	118.733	0.000
$\ln$(Market Capitalisation)	-0.0021***	1.95e-05	-107.393	0.000
Popularity Weight	0.7071***	0.017	41.952	0.000
Volatility Lag -1	0.3854***	0.003	136.482	0.000
Volatility Lag -2	0.1800***	0.003	71.002	0.000
Volatility Lag -3	0.1635***	0.002	71.635	0.000
No. Observations	1718133			
R-squared	0.456			
Adj. R-squared	0.456			

*** denotes significance at the 1 per cent level.

The lagged model has a higher R-squared value, and the result is more robust as the Durbin-Watson test result of 2 indicates no autocorrelation. Additionally, this regression suggests that past volatility levels influence current volatility, as each lagged volatility variable exhibits a positive relationship with the dependent variable. The results are economically significant as for every 1% (0.01 unit) increase in weight in the Robinhood user's total portfolios (popularity weight), the volatility is estimated to increase by approximately 0.71% (0.007106 units), all else being equal. Results suggest that small changes in individual ownership variables can have a notable impact on the volatility of a stock in the context of the regression model. Concerning the first hypothesis, and based on these results, we can reject the null hypothesis that company size and the level of individual investor ownership do not significantly impact volatility.

The first set of regression examined the impact of popularity weight and market capitalisation on volatility. The findings from the two regressions align well with the prior studies investigating the relationship between individual investor behaviours and stock price volatility. Firstly, the positive association between popularity weight and volatility in the regression model corresponds to the findings of Baig et al. (2023), Aharon et al. (2022), Kyröläinen (2008) and Foucault et al. (2011), who all highlighted the positive impact of retail trading on volatility. Interestingly, the findings align with those of institutional investors by Sias (1996). Thus, investor interest, regardless of investor group, is associated with higher levels of volatility.

4.3.2 Changes in volatility and Robinhood holdings

The contemporaneous relationship between individual investors and volatility found in Table 6, Table 7, and Table 8 is consistent with the two stories. First, securities that exhibit greater volatility may attract individual investor investors. Secondly, an increase in individual investor holdings induces an increase in volatility. These conclusions confirm the suggestions by Kumar (2009) that individual investors may be attracted to more volatile stocks.

To test whether more volatile securities attract individual investors, we will examine whether lagged change popularity weight can explain volatility beyond what is explained by lag volatility. The results from table Table 9 indicate that a change in volatility is positively associated with the level of individual investor ownership, suggesting that investors may adjust their portfolios in response to changes in volatility. These results indicate a highly significant regression model with an R-squared value of 1.00. The result is significant at the p = 0.01 level. Concerning the second hypothesis, and based on the results, we can reject the null hypothesis that the change in volatility does not significantly influence the level of individual investor ownership.

The popularity weight variable is explained mainly by the popularity weight of previous trading days. The change in volatility is statistically significant, but the economic significance is more challenging to decipher. Based on the results, holding all else equal, if the stock volatility doubled, individual investor ownership of stock in the Robinhood portfolio would increase by 0.22%.

Table 9. Change in volatility and changes in individual investor holdings.

Variable	Coefficients	Std. Error	z	P-value
Intercept	-6.021e-07***	4.76e-08	-12.655	0.000
Change Volatility Lag -1	1.399e-05***	1.02e-06	13.721	0.000
Popularity Lag -1	0.9993***	0.000	8886.476	0.000
Return Lag -1	1.167e-05***	1.7e-06	6.881	0.000
No. Observations	1616023			
R-squared	1.000			
Adj. R-squared	1.000			

*** denotes significance at the 1 per cent level.

Comparing results to the earlier findings by Sias (1996), we can demonstrate several differences between individual and institutional investors. For individual investors, there is a positive relationship between a positive change in volatility and individual investor holdings, indicating that an increase in volatility attracts individual investors. Findings differ from those of Sias (1996), who finds only limited evidence of institutional investors being attracted to securities with increased volatility. Moreover, Sias (1996) finds that institutional investors are more likely to get rid of their stocks if volatility increases.

4.3.3 Changes in Robinhood holdings and volatility

To test whether an increase in individual investor holdings is associated with a subsequent increase in volatility, we examined whether the stock's current volatility is a function of past volatility, change in popularity weight, and past returns. Three models were tested. Models differed in terms of lags for changes in popularity, ranging from one to three days. Model (1) is a model with a lag of one day, model (2) with two days and model (3) with three days for popularity weight, while lags for volatility and returns remain one day.

Table 10. Change in individual investor holdings and volatility.

	Estimated models		
	(1)	(2)	(3)
Intercept	0.0178***	0.0178***	0.0178***
	(0.000)	(0.001)	(0.001)
Volatility	0.6257***	0.6257***	0.6259***
	(0.002)	(0.002)	(0.002)
Change in Popularity weight	0.2060	13.0058***	24.3103***
	(7.214)	(5.040)	(5.277)
Return	-0.0271***	-0.0270***	-0.0268***
	(0.003)	(0.003)	(0.003)
Observations	1674204	1666909	1655981
R-squared	0.389	0.389	0.389
Adj. R-squared	0.389	0.389	0.389

***p<0.01, **p<0.05, *p<0.1. Standard errors are in parentheses.

Table 10 presents results from three estimated models examining the relationship between changes in individual investor holdings and volatility. For volatility, the results are significant for models (2) and (3). Thus, the results show that lag changes in individual investor ownership predict variation in volatility in two—and three-day lags. Results are not significant at one-day lags. Concerning the third hypothesis, and based on these results, we still reject the null hypothesis that changes in individual investor ownership do not predict volatility.

It is interesting to note that the coefficient associated with the lag return is negative and statistically significant. Negative returns are associated with higher levels of volatility. On the other hand, the results suggest that higher returns result in relatively lower volatility. Avramov et al. (2006) findings associated with non-informative trades for herding investors suggest that after a stock price drop, there is an increase in volatility. Thus, we could assume, based on the results, that Robinhood traders are noise traders.

4.4 Discussion

The thesis results can be summarised in three points. First, controlling for capitalisation, higher levels of volatility are associated with higher levels of individual investor ownership. Second, a change in volatility is positively associated with the level of individual investor ownership. Third, individual investor ownership predicts variation in volatility with a lag of two and three days. Thus, it is concluded that individual investor interest is associated with volatility. Furthermore, the findings broadly support the work of other studies linking individual investors and volatility.

The first finding regarding the contemporaneous relationship between volatility and individual investors is significant in two non-parametric methods and two regressions. Significant differences were observed between deciles based on individual investor ownership and volatility in the Kruskal-Wallis test. Additionally, using the Wilcoxon rank-sum test, higher volatilities were observed in the high popularity weight groups, supporting the idea that individual investors are drawn to stocks with greater volatility. Later, two regressions confirm that independent of company size, higher levels of volatility are associated with higher levels of individual investor ownership. Thus, this study finds evidence that supports findings from previous studies, e.g. Baig et al. (2023), Aharon et al. (2022), Foucault et al. (2011), and Kyröläinen (2008).

The second finding is that a change in volatility is positively associated with increased individual investor ownership. This association is both statistically and economically significant. This result supports the idea that individual investors are attracted to stocks with greater volatility Kumar (2009) discovered. Moreover, the result is consistent with the findings of Aharon et al. (2022). Aharon et al. (2022) found that the growth rate of Robinhood accounts holding the stock in American depositary receipts is correlated with increased volatility. More broadly, the result from the third regression corroborates the ideas of Baig et al. (2023), who suggested that individual trading leads to increased volatility levels, especially during crisis periods. Conversely, Foucault et al. (2011) suggest that individual investors' reduced interest and lower participation in trading led to decreased volatility. Our thesis finds the reverse, indicating that increased interest measured by popularity weight leads to increased volatility.

The third finding is that individual investor ownership predicts variation in volatility with a lag of two and three days. Although the regression with one-day lags for changes in individual investor holdings is insignificant, it indicates the same directional effect. One possible explanation, supported by conclusions from extreme changes tables, is that increases in Robinhood holdings may lead to volatility with a delay. Additionally, negative returns are associated with higher levels of volatility. This behaviour might be indicative of either contrarian or herding behaviour. The positive associated with lagged volatilities were likely due to contrarian trading of individual investors as extreme increases in popularity weight followed significant negative returns. Avramov et al. (2006) discovered that non-informed trades associated with herding investors increase volatility after a stock price decline. Thus, the results from the regression concerning changes in volatility and Robinhood holdings show that individual investors exhibit behaviour associated with herding investors, as evidenced by a relationship between increasing volatility and negative returns. This finding illustrates the conclusion Raafat et al. (2009) put forward that herding includes theoretical and methodical approaches that are diverse and disconnected.

The results from two non-parametric methods reveal patterns of individual investor behaviour observed in previous studies. Generally, Robinhood investors prefer both extremely low-volatility and high-volatility stocks. Additionally, Robinhood investors prefer smaller companies' stocks in smaller market capitalisation deciles and larger companies in larger market capitalisation deciles. Although higher volatility is preferred across market capitalisations, the two highest market capitalisation deciles with the lowest volatilities have the highest median popularity weights. Additionally, Robinhood investors seem to prefer stocks that resemble lotteries with high volatilities, aligning with findings on individual investor preferences by Kumar (2009).

On the other hand, the preference for low-volatility stocks could be explained by risk aversion during periods of increased ambiguity, as discovered by Kostopoulos et al. (2022). An alternative explanation could be the two distinct behavioural patterns among gamblers identified by Weiss-Cohen et al. (2024). During low volatility environments, gamblers exhibited more investment-like behaviours with fewer trades, while during high volatility, their actions were more akin to gambling, characterised by increased trading activity.

Robinhood users' portfolios have, on average, about three stock positions. This tendency for investors to maintain underdiversified or, in other words, concentrated portfolios has been highlighted by Dorn and Huberman (2005). Additionally, the volatility preferences observable from non-parametric test results could be explained by how individual investors understand the trade-off between skewness and diversification (Mitton & Vorkink, 2007). Thus, high volatility could be seen as a heuristic for higher returns.

Compared to other studies examining the relationship between individual investors and volatility, this thesis offers more direct evidence of individual investors being drawn to more volatile stocks. By examining individual investor ownership of stocks, we can compare our results with past studies concerning institutional investor ownership. Our thesis identifies similarities between individual investors and institutional investors. Following the findings by Sias (1996) regarding institutional investors, this study indicates that individual investors are associated with higher levels of volatility when capitalisation is held constant. This relationship is found in both non-parametric tests and is statistically significant regressions. We could conclude, more broadly, that increased investor interest is associated with higher volatility levels.

The Kruskal-Wallis test confirms statistically significant differences in popularity weight and volatility across various market capitalisation deciles. Consistent with past studies, market capitalisation and volatility demonstrate an inverse relationship. Additionally, the results reveal a connection between popularity weight and market capitalisation. However, a deviation observed in the smallest deciles suggests that individual investor volatility preferences may differ from institutional investor preferences, as observed in Sias (1996). Additional differences are found to exist between institutional investors and individual investors. Sias (1996) suggests that institutional investors are not notably attracted to increased volatility. Instead, they tend to sell volatile securities from their portfolios. In contrast, our thesis finds that for individual investors, there is a positive relationship between an increase in volatility and individual investor holdings, indicating that an increase in volatility attracts individual investors.

In terms of methodology, there are key differences from prior studies that should be discussed. This thesis adopts the methodology of Sias (1996). While Sias (1996) focused on institutional investors and implicitly considered individual investors, this thesis focuses on individual investors only. An additional difference is the time frame. Thus, as Sias (1996) finds that changes in institutional holdings predict volatility yearly, our study suggests that changes in individual investor holdings over two to three days also forecast volatility.

Additionally, this measure of individual interest significantly differs from the approach of Aharon et al. (2022), where two variables were utilised separately to study the level and growth of Robinhood holdings. While Aharon et al. (2022) employ measures to investigate the impact of trading, the popularity weight used in this thesis facilitates the examination of the impact of individual investor ownership. The thesis' measure of individual investor ownership is also derived from the Robintrack dataset. The 'popularity weight' variable is based on the novel measure introduced by Welch (2022). This measure resembles the percentage holding in a household's equally-weighted portfolio. The popularity weight variable considers the count of unique accounts holding at least one share of a particular ordinary stock, adjusted by the total number of accounts holding at least one share of any asset on Robinhood.

The thesis findings reinforce the conclusions drawn by Aharon et al. (2022) regarding the influence of Robinhood traders and the policy implications concerning traders' impact on financial markets. Aharon et al. (2022) study indicates that increased access to financial markets for individual investors may result in heightened volatility, particularly evident in American depositary receipts. This thesis extends this conclusion to the broader spectrum of ordinary shares accessible via Robinhood. While individual investors contribute to market liquidity, this liquidity comes with the cost of increased volatility.

5 Conclusion

This study set out to establish how individual investor ownership impacts or may be influenced by volatility. We achieved this aim by utilising non-parametric methods and linear regressions to study a novel dataset that includes information on individual investor ownership from the Robinhood trading platform for a sample of 3765 ordinary stocks. While previous studies using the Robintrack dataset have explored different aspects of Robinhood users' behaviour, none have directly studied the relationship between Robinhood user's ownership of ordinary stocks and volatility.

This study identified that individual investors are attracted to increased volatility. The results also showed that individual investor interest is associated with increased volatility, regardless of market capitalisation. Thus, the study's findings on individual investors' relationship with volatility are twofold. First, highly volatile securities may attract individual investors. Second, an increase in individual investor holdings can lead to higher volatility. These findings are consistent with the conclusions of past studies on individual investor behaviour and the relationship between individual investors and volatility. Additionally, the tendency for investors to maintain under-diversified portfolios could explain volatility as Robinhood investors, on average, own concentrated portfolios. Furthermore, the tendency towards risk aversion during uncertain periods may explain observed patterns of highest ownership percentages in both low and high-volatility stocks.

The thesis aimed to follow the approach outlined in Sias (1996), which, in turn, allowed us to compare institutional and individual investor ownership. The findings mostly align. However, while institutional investors tend to reduce holdings in volatile securities, individual investors are attracted to increased volatility. The comparability of the findings may be limited by differences in time frames and metrics related to ownership. Additionally, the sample may not fully represent the population of individual investors as not all individual investors are Robinhood users, and not every Robinhood user is necessarily an individual investor. However, given the law of large numbers, the Robintrack dataset should be a representative sample of individual investors.

A natural progression of this thesis would be to analyse individual investor and institutional ownership data simultaneously. Additionally, this could expand the understanding of how the two investor groups interact in the markets. Institutional investors, including hedge funds, used Robintrack to monitor small investors and developed rule-based frameworks using the Robintrack data (Basek & Ponczek, 2020). This monitoring indicates growing interest from institutional investors in retail activity. Further research could also be conducted to determine the effectiveness of extending the methodology by Bushee (1998) to be used with individual investor ownership data. Bushee (1998) classifies institutional investors as dedicated institutions, quasi-indexers, and transient institutions. Incorporating institutional investors' preferences by categorising stocks with this classification could be used to clarify the relationship between individual and institutional investors. A similar methodology could extended to be used with individual investors. Stocks could be categorised by the turnover rate in the consensus Robinhood portfolio. Thus, stocks could be classified as popular among traders or passive investors, allowing us to classify types of individual investors.

Bibliography

Abreu, D., & Brunnermeier, M. K. (2002). Synchronization risk and delayed arbitrage. *Journal of Financial Economics*, *66*(2-3), 341–360. (Publisher: Elsevier)

Aharon, D. Y., Baig, A. S., & Delisle, R. J. (2022). The impact of Robinhood traders on the volatility of cross-listed securities. *Research in International Business and Finance*, *60*, 101619. (Publisher: Elsevier)

Akerlof, G., & Shiller, R. (2009). *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*. Princeton University Press. (tex.lccn: 2008052649) Retrieved from https://books.google.fi/books?id=wl_GF81cSvUC

Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, *57*(3), 1047–1091. (Publisher: Wiley Online Library)

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, 885–905. (Publisher: JSTOR)

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of financial economics*, *61*(1), 43–76. (Publisher: Elsevier)

Avery, C., & Zemsky, P. (1998). Multidimensional uncertainty and herd behavior in financial markets. *American economic review*, 724–748. (Publisher: JSTOR)

Avramov, D., Chordia, T., & Goyal, A. (2006). The impact of trades on daily volatility. *The Review of Financial Studies*, *19*(4), 1241–1277. (Publisher: Oxford University Press)

Baig, A. S., Blau, B. M., Butt, H. A., & Yasin, A. (2023). Reprint of: Do retail traders destabilize financial markets? An investigation surrounding the COVID-19 pandemic. *Journal of banking & finance*, *147*, 106744. (Publisher: Elsevier)

Balasubramaniam, V., Campbell, J. Y., Ramadorai, T., & Ranish, B. (2023). Who owns what? A factor model for direct stockholding. *The Journal of Finance*, *78*(3), 1545–1591. (Publisher: Wiley Online Library)

Banerjee, A. V. (1992). A simple model of herd behavior. *The quarterly journal of eco-*

nomics, *107*(3), 797–817. (Publisher: MIT Press)

Barber, B. M., Huang, X., Odean, T., & Schwarz, C. (2022, December). Attention-Induced Trading and Returns: Evidence from Robinhood Users. *The Journal of Finance*, *77*(6), 3141–3190. https://doi.org/10.1111/jofi.13183

Barber, B. M., Lee, Y.-T., Liu, Y.-J., & Odean, T. (2009). Just how much do individual investors lose by trading? *The Review of Financial Studies*, *22*(2), 609–632. (Publisher: Oxford University Press)

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The journal of Finance*, *55*(2), 773–806. (Publisher: Wiley Online Library)

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, *21*(2), 785–818. (Publisher: Oxford University Press)

Barber, B. M., & Odean, T. (2013). The behavior of individual investors. In *Handbook of the economics of finance* (Vol. 2, pp. 1533–1570). Elsevier.

Barber, B. M., Odean, T., & Zhu, N. (2008). Do retail trades move markets? *The Review of Financial Studies*, *22*(1), 151–186. (Publisher: Society for Financial Studies)

Barber, B. M., Odean, T., & Zhu, N. (2009, January). Do Retail Trades Move Markets? *Review of Financial Studies*, *22*(1), 151–186. https://doi.org/10.1093/rfs/hhn035

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, *49*(3), 307–343. https://doi.org/10.1016/S0304-405X(98)00027-0

Basek, S., & Ponczek, S. (2020, July). Hedge Funds Approach Robintrack to Keep Eyes on Tiny Investors. *Bloomberg.com*. Retrieved from [2024-04-14]https://www.bloomberg.com/news/articles/2020-07-23/hedge-funds-approach-robintrack-to-keep-eyes-on-tiny-investors

Beckers, S. (1983). Variances of Security Price Returns Based on High, Low, and Closing Prices. *The Journal of Business*, *56*(1), 97–112. Retrieved from http://www.jstor.org/stable/2352748

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, *100*(5), 992–1026. (Publisher: The University of Chicago Press)

Black, F. (1986, July). Noise. *The Journal of Finance*, *41*(3), 528–543. https://doi.org/10.1111/j.1540-6261.1986.tb04513.x

Blasco, N., Corredor, P., & Ferreruela, S. (2012). Does herding affect volatility? Implications for the Spanish stock market. *Quantitative Finance*, *12*(2), 311–327. (Publisher: Taylor & Francis)

Brandt, M. W., Brav, A., Graham, J. R., & Kumar, A. (2010). The idiosyncratic volatility puzzle: Time trend or speculative episodes? *The Review of Financial Studies*, *23*(2), 863–899. (Publisher: Society for Financial Studies)

Brooks, C. (2014). *Introductory econometrics for finance* (Third edition ed.). Cambridge ; New York: Cambridge University Press.

Bryzgalova, S., Pavlova, A., & Sikorskaya, T. (2023). Retail trading in options and the rise of the big three wholesalers. *The Journal of Finance*, *78*(6), 3465–3514. (Publisher: Wiley Online Library)

Bushee, B. J. (1998). The influence of institutional investors on myopic R&D investment behavior. *Accounting review*, 305–333. (Publisher: JSTOR)

Bushee, B. J. (2001). Do institutional investors prefer near-term earnings over long-run value? *Contemporary accounting research*, *18*(2), 207–246. (Publisher: Wiley Online Library)

Bushee, B. J., & Noe, C. F. (2000). Corporate disclosure practices, institutional investors, and stock return volatility. *Journal of accounting research*, 171–202. (Publisher: JSTOR)

Campbell, J. Y., & Vuolteenaho, T. (2004). Bad beta, good beta. *American Economic Review*, *94*(5), 1249–1275. (Publisher: American Economic Association)

Chan, K., & Fong, W.-M. (2000). Trade size, order imbalance, and the volatility–volume relation. *Journal of Financial Economics*, *57*(2), 247–273. (Publisher: Elsevier)

Chang, E. C., Cheng, J. W., & Khorana, A. (2000). An examination of herd behavior in equity markets: An international perspective. *Journal of Banking & Finance*, *24*(10), 1651–1679. (Publisher: Elsevier)

Choi, N., & Sias, R. W. (2009). Institutional industry herding. *Journal of financial economics*, *94*(3), 469–491. (Publisher: Elsevier)

Christie, W. G., & Huang, R. D. (1995). Following the pied piper: do individual returns herd around the market? *Financial Analysts Journal*, *51*(4), 31–37. (Publisher: Taylor & Francis)

Cipriani, M., & Guarino, A. (2005). Herd behavior in a laboratory financial market. *American Economic Review*, *95*(5), 1427–1443. (Publisher: American Economic Association)

Cole, C. R. (2014). Volatility: The market price of uncertainty. In *CFA Institute Conference Proceedings Quarterly* (Vol. 31, pp. 36–47). (Issue: 1)

Daniel, K. D., Hirshleifer, D. A., & Subrahmanyam, A. (1997). A theory of overconfidence, self-attribution, and security market under-and over-reactions. *Self-Attribution, and Security Market Under-and Over-Reactions (February 19, 1997)*.

Daniel, W. W. (1978). *Applied nonparametric statistics*. Boston: Houghton Mifflin.

De Bondt, W. F., & Forbes, W. P. (1999). Herding in analyst earnings forecasts: evidence from the United Kingdom. *European financial management*, *5*(2), 143–163. (Publisher: Wiley Online Library)

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990a). Noise Trader Risk in Financial Markets. *Journal of Political Economy*, *98*(4), 703–738. (Publisher: University of Chicago Press) Retrieved from [2022-10-16]https://www.jstor.org/stable/2937765

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990b). Positive feedback investment strategies and destabilizing rational speculation. *the Journal of Finance*, *45*(2), 379–395. (Publisher: Wiley Online Library)

Dennis, P. J., & Strickland, D. (2002). Who blinks in volatile markets, individuals or institutions? *The Journal of Finance*, *57*(5), 1923–1949. (Publisher: Wiley Online Library)

Dorn, D., & Huberman, G. (2005). Talk and action: What individual investors say and what they do. *Review of Finance*, *9*(4), 437–481. (Publisher: Oxford University Press)

Dorn, D., & Huberman, G. (2010). Preferred risk habitat of individual investors. *Journal of Financial Economics*, *97*(1), 155–173. (Publisher: Elsevier)

Drehmann, M., Oechssler, J., & Roider, A. (2005). Herding and contrarian behavior in financial markets: An internet experiment. *American Economic Review*, *95*(5), 1403–

1426. (Publisher: American Economic Association)

Eaton, G. W., Green, T. C., Roseman, B., & Wu, Y. (2021). Zero-commission individual investors, high frequency traders, and stock market quality. *High Frequency Traders, and Stock Market Quality (January 2021)*.

Eaton, G. W., Green, T. C., Roseman, B. S., & Wu, Y. (2022). Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics*, *146*(2), 502–528. (Publisher: Elsevier)

Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*, *38*(1), 34–105. (Publisher: University of Chicago Press) Retrieved from [2021-03-28]https://www.jstor.org/stable/2350752

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383–417. (Publisher: [American Finance Association, Wiley]) https://doi.org/10.2307/2325486

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, *33*(1), 3–56. (Publisher: Elsevier)

Fama, E. F., & French, K. R. (2015, April). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010

Foucault, T., Sraer, D., & Thesmar, D. J. (2011, August). Individual Investors and Volatility. *The Journal of Finance*, *66*(4), 1369–1406. https://doi.org/10.1111/j.1540-6261.2011.01668.x

French, K. R., Schwert, G. W., & Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of financial Economics*, *19*(1), 3–29. (Publisher: Elsevier)

Friedman, H. L., & Zeng, Z. (2022). Retail investor trading and market reactions to earnings announcements. *Available at SSRN 3817979*.

Gallant, A. R., Hsu, C.-T., & Tauchen, G. (1999, November). Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance. *Review of Economics and Statistics*, *81*(4), 617–631. https://doi.org/10.1162/003465399558481

Garman, M. B., & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of business*, 67–78. (Publisher: JSTOR)

Gleason, K. C., Mathur, I., & Peterson, M. A. (2004). Analysis of intraday herding behavior among the sector ETFs. *Journal of empirical Finance*, *11*(5), 681–694. (Publisher: Elsevier)

Greenwood, R., & Thesmar, D. (2011). Stock price fragility. *Journal of Financial Economics*, *102*(3), 471–490. (Publisher: Elsevier)

Grinblatt, M., & Keloharju, M. (2000). The investment behavior and performance of various investor types: a study of Finland's unique data set. *Journal of financial economics*, *55*(1), 43–67. (Publisher: Elsevier)

Grinblatt, M., & Keloharju, M. (2001, April). What Makes Investors Trade? *The Journal of Finance*, *56*(2), 589–616. https://doi.org/10.1111/0022-1082.00338

Grinblatt, M., Titman, S., & Wermers, R. (1995). Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. *The American economic review*, 1088–1105. (Publisher: JSTOR)

Gutierrez Jr, R. C., & Kelley, E. K. (2008). Institutional herding: Destabilizing buys, stabilizing sells. *CFA Digest*, *38*, 39–40.

Hirshleifer, D., & Hong Teoh, S. (2003). Herd behaviour and cascading in capital markets: A review and synthesis. *European Financial Management*, *9*(1), 25–66. (Publisher: Wiley Online Library)

Hong, H., & Stein, J. C. (1999). A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets. *The Journal of Finance*, *54*(6), 2143–2184. (Publisher: [American Finance Association, Wiley]) Retrieved from [2022-10-16]https://www.jstor.org/stable/797990

Hwang, S., & Salmon, M. (2004). Market stress and herding. *Journal of Empirical Finance*, *11*(4), 585–616. (Publisher: Elsevier)

Jones, C. M., Kaul, G., & Lipson, M. L. (1994). Transactions, volume, and volatility. *The Review of Financial Studies*, *7*(4), 631–651. (Publisher: Oxford University Press)

Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I* (pp. 99–127). World Scientific.

Kelley, E. K., & Tetlock, P. C. (2013). How wise are crowds? Insights from retail orders and stock returns. *The Journal of Finance*, *68*(3), 1229–1265. (Publisher: Wiley Online Library)

Kendall, C. (2023). Herding and contrarianism: A matter of preference? *Review of Economics and Statistics*, *105*(1), 190–205. (Publisher: MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . .)

Keynes, J. M. (1936). *The general theory of employment, interest and money*. Macmillan. (tex.added-at: 2009-06-26T15:25:19.000+0200 tex.description: diverse cognitive systems bib tex.interhash: 060c8b5ac81500224415ba44501f4164 tex.intrahash: c27c51d717b88bdf405c38cbfdfa65ac tex.timestamp: 2009-06-26T15:25:41.000+0200)

Kim, W., & Wei, S.-J. (2002). Foreign portfolio investors before and during a crisis. *Journal of international economics*, *56*(1), 77–96. (Publisher: Elsevier)

Kostopoulos, D., Meyer, S., & Uhr, C. (2022, July). Ambiguity about volatility and investor behavior. *Journal of Financial Economics*, *145*(1), 277–296. https://doi.org/10.1016/j.jfineco.2021.07.004

Kumar, A. (2009). Who gambles in the stock market? *The journal of finance*, *64*(4), 1889–1933. (Publisher: Wiley Online Library)

Kumar, A., & Lee, C. M. (2006). Retail investor sentiment and return comovements. *The Journal of Finance*, *61*(5), 2451–2486.

Kyröläinen, P. (2008). Day trading and stock price volatility. *Journal of Economics and Finance*, *32*, 75–89. (Publisher: Springer)

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1992). The impact of institutional trading on stock prices. *Journal of financial economics*, *32*(1), 23–43. (Publisher: Elsevier)

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *The journal of finance*, *49*(5), 1541–1578. (Publisher: Wiley Online Library)

Li, W., Rhee, G., & Wang, S. S. (2017). Differences in herding: Individual vs. institutional investors. *Pacific-Basin Finance Journal*, *45*, 174–185. (Publisher: Elsevier)

Liew, J., & Vassalou, M. (2000). Can book-to-market, size and momentum be risk factors that predict economic growth? *Journal of Financial Economics*, *57*(2), 221–245. (Publisher: Elsevier)

Luo, C., Ravina, E., Sammon, M., & Viceira, L. (2022). Retail investors' contrarian behavior around news. *Attention*, *and the Momentum Effect*.

Malkiel, B. G. (1989). Efficient Market Hypothesis. In J. Eatwell, M. Milgate, & P. Newman

(Eds.), *Finance* (pp. 127–134). London: Palgrave Macmillan UK.

Marcus, E. (2020, December). 'You Checked Tesla the Most': Robinhood Recaps From a Volatile Year. *The New York Times*. Retrieved from [2024-04-05]https://www.nytimes.com/2020/12/24/style/robinhood-2020-recaps.html

Markowitz, H. (1952, March). Portfolio Selection. *The Journal of Finance*, *7*(1), 77. https://doi.org/10.2307/2975974

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of financial economics*, *8*(4), 323–361. (Publisher: Elsevier)

Mitton, T., & Vorkink, K. (2007). Equilibrium underdiversification and the preference for skewness. *The Review of Financial Studies*, *20*(4), 1255–1288. (Publisher: Oxford University Press)

Moss, A., Naughton, J. P., & Wang, C. (2023). The irrelevance of environmental, social, and governance disclosure to retail investors. *Management Science*. (Publisher: INFORMS)

Newey, W. K., & West, K. D. (1987, May). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3), 703. https://doi.org/10.2307/1913610

Nofsinger, J. R., & Sias, R. W. (1999). Herding and feedback trading by institutional and individual investors. *The Journal of finance*, *54*(6), 2263–2295. (Publisher: Wiley Online Library)

Odean, T. (1998). Are investors reluctant to realize their losses? *The Journal of finance*, *53*(5), 1775–1798. (Publisher: Wiley Online Library)

Odean, T. (1999). Do investors trade too much? *American economic review*, *89*(5), 1279–1298. (Publisher: American Economic Association)

Ozik, G., Sadka, R., & Shen, S. (2021). Flattening the illiquidity curve: Retail trading during the COVID-19 lockdown. *Journal of Financial and Quantitative Analysis*, *56*(7), 2356–2388. (Publisher: Cambridge University Press)

Pagano, M. S., Sedunov, J., & Velthuis, R. (2021). How did retail investors respond to the COVID-19 pandemic? The effect of Robinhood brokerage customers on market quality. *Finance Research Letters*, *43*, 101946. (Publisher: Elsevier)

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of business*, 61–65. (Publisher: JSTOR)

Raafat, R. M., Chater, N., & Frith, C. (2009). Corrigendum: Herding in humans. *Trends in Cognitive Sciences*, *13*(12), 504. (Publisher: Elsevier)

Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (Eds.). (1998). *Applied Regression Analysis*. New York: Springer-Verlag. https://doi.org/10.1007/b98890

Rogers, L. C. G., & Satchell, S. E. (1991). Estimating variance from high, low and closing prices. *The Annals of Applied Probability*, 504–512. (Publisher: JSTOR)

Schlarbaum, G. G., Lewellen, W. G., & Lease, R. C. (1978a). The common-stock-portfolio performance record of individual investors: 1964–70. *The Journal of Finance*, *33*(2), 429–441. (Publisher: Wiley Online Library)

Schlarbaum, G. G., Lewellen, W. G., & Lease, R. C. (1978b). Realized returns on common stock investments: The experience of individual investors. *Journal of Business*, 299–325. (Publisher: JSTOR)

Schwert, G. W. (1989). Why does stock market volatility change over time? *The journal of finance*, *44*(5), 1115–1153. (Publisher: Wiley Online Library)

Schwert, G. W. (1990). Stock volatility and the crash of'87. *The review of financial studies*, *3*(1), 77–102. (Publisher: Oxford University Press)

Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In *9th python in science conference.*

Shefrin, H., & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of finance*, *40*(3), 777–790. (Publisher: Wiley Online Library)

Shiller, R. J. (2002). Bubbles, human judgment, and expert opinion. *Financial Analysts Journal*, *58*(3), 18–26. (Publisher: Taylor & Francis)

Shiller, R. J. (2003). From Efficient Markets Theory to Behavioral Finance. *Journal of Economic Perspectives*, *17*(1), 83–104. https://doi.org/10.1257/089533003321164967

Shiller, R. J. (2015). *Irrational exuberance* (Revised and expanded third edition ed.). Princeton: Princeton University Press.

Shleifer, A., & Summers, L. H. (1990). The noise trader approach to finance. *Journal of Economic perspectives*, *4*(2), 19–33. (Publisher: American Economic Association)

Sias, R. W. (1996). Volatility and the institutional investor. *Financial Analysts Journal*, *52*(2), 13–20. (Publisher: Taylor & Francis)

Sias, R. W. (2004). Institutional herding. *The Review of Financial Studies*, *17*(1), 165–206. (Publisher: Oxford University Press)

Spyrou, S. (2013). Herding in financial markets: a review of the literature. *Review of Behavioral Finance*, *5*(2), 175–194. (Publisher: Emerald Group Publishing Limited)

Tan, G. K. S. (2021). Democratizing finance with Robinhood: Financial infrastructure, interface design and platform capitalism. *Environment and Planning A: Economy and Space*, *53*(8), 1862–1878. (Publisher: SAGE Publications Sage UK: London, England)

Tan, L., Chiang, T. C., Mason, J. R., & Nelling, E. (2008). Herding behavior in Chinese stock markets: An examination of A and B shares. *Pacific-Basin finance journal*, *16*(1-2), 61–77. (Publisher: Elsevier)

Thaler, R. H., & Barberis, N. (2002). *A survey of behavioral finance*. National Bureau of Economic Research.

Turner, P. (2020). Critical values for the Durbin-Watson test in large samples. *Applied Economics Letters*, *27*(18), 1495–1499. (Publisher: Taylor & Francis)

Venkatachalam, M. (2000). Discussion of corporate disclosure practices, institutional investors, and stock return volatility. *Journal of Accounting Research*, *38*, 203–207. (Publisher: JSTOR)

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., … SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272.

Weiss-Cohen, L., Newall, P., Bart, Y., Zloteanu, M., Peacey, M., Ayton, P., & Clacher, I. (2024). When vegas comes to wall street: Associations between stock price volatility and trading frequency amongst gamblers. *International Journal of Mental Health and Addiction*, 1–20. (Publisher: Springer)

Welch, I. (2000). Herding among security analysts. *Journal of Financial economics*, *58*(3), 369–396. (Publisher: Elsevier)

Welch, I. (2022, June). The Wisdom of the Robinhood Crowd. *The Journal of Finance*, *77*(3), 1489–1527. https://doi.org/10.1111/jofi.13128

Wermers, R. (1999). Mutual Fund Herding and the Impact on Stock Prices. *The Journal of Finance*, *54*(2), 581–622. (_eprint:

https://onlinelibrary.wiley.com/doi/pdf/10.1111/0022-1082.00118)
https://doi.org/10.1111/0022-1082.00118

Wes McKinney. (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt & Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (p. 56 - 61).

White, H. (1980, May). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, *48*(4), 817. https://doi.org/10.2307/1912934

Wooldridge, J. M. (2013). *Introductory econometrics: a modern approach* (5th ed ed.). Mason, Ohio: South-Western, Cengage Learning.

World Health Organization. (2020, April). *Archived: WHO Timeline - COVID-19*. Retrieved from [2024-04-16]https://www.who.int/news/item/27-04-2020-who-timeline---covid-19

Yang, D., & Zhang, Q. (2000). Drift-independent volatility estimation based on high, low, open, and close prices. *The Journal of Business*, *73*(3), 477–492. (Publisher: JSTOR)