Libera Università Internazionale degli Studi Sociali Luiss Guido Carli

Degree Program in *Economia e Management*

Course of *Financial Market Analysis*

Bachelor's Degree Thesis

THE EFFECTS OF SOCIAL NETWORKS AND FINFLUENCERS ON FINANCIAL MARKET VOLATILITY, LIQUIDITY AND INVESTORS' DECISION-MAKING

Supervisor
Prof. Giovanni Rillo

Candidate
Matteo De Thomasis
285761

Academic Year 2024/2025

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.INTRODUCTION

1.1 Thesis objective

In recent years, social networks have become increasingly popular among the vast majority of people and have also allegedly become capable of shaping people's views, ideas and opinions. This rapid digital transformation has raised numerous ethical issues and practical problems that were not even imaginable a few years ago. Socials such as YouTube, Twitter, Reddit, TikTok and Instagram have emerged as an immense source of ideas and have consequently acquired a significant role in influencing opinions. Given that the content on social media is endless and varied, it is only natural that discussions also involve the financial world.

An average, prudent investor, who 50 years ago chose how to invest his savings based on the analysis of traditional media, which have many limitations, can now easily acquire thousands of information and opinions simply by having an Internet connection. The simplicity and democratisation of market access has brought millions of new investors into the financial ecosystem, which is now much less elitist. In an age where financial advice and opinions are increasingly shared through short videos and anonymous posts, understanding the dynamics of online influence is not only timely but necessary for regulators, market analysts and investors.

Considering these transformations, it was foreseeable that the opinion of some individuals would have a significant impact on others, leading to the emergence of full-time financial influencers, the so-called "finfluencers", who play a key role in the dissemination of financial advice and potentially impact market dynamics. These influencers combine entertainment and advice and, in so doing, incorporate an emotional aspect into financial advice, potentially inducing their followers to act irresponsibly.

The consequences of this enormous change are mostly visible, but it would be pretentious to assume that all the upheavals resulting from this revolution are only understandable at first glance. For this reason, the aim of this thesis is to understand how social networks and finfluencers impact the decision-making process, volatility and liquidity of the financial markets: thus: given the subject matter, this paper will follow a long process of studies on the subject.

The dissertation analyses how various research studies on the subject are challenging traditional market theories such as the Efficient Market Hypothesis, furthermore, this study aims to present irrational consumer behaviour, such as herding or overconfidence, caused by online communities and finfluencers. To achieve the thesis objective of better understanding the relationship between social networks and the financial market, this paper uses both a qualitative and quantitative approach.

Qualitatively, the study analyses a wide range of existing literature focusing on behavioural finance, the role of emotions in investor psychology and psychophysical approaches related to behavioural finance. The aim is to inform the readers about the state of the art and make them aware of the most important findings that could be useful for achieving the objective of this thesis.

At a quantitative level, instead, the paper conducts both an analysis based on the notorious Reddit-GameStop case and one on the Elon-Dogecoin case. The choice to conduct the first study using Reddit is both methodological and practical: the latter provides easily obtainable open access information, a useful feature especially after the recent policy changes that severely restrict the use of Twitter/X API. The analysis takes a confrontational approach and aims to compare the liquidity and volatility of financial markets with the sentiment analysis of a particular subreddit over a short period of time. This approach makes it possible to directly investigate potential correlations between the market and social media, which can be detected with basic statistical models. While the second analysis on finfluencers (specifically, the Elon-Dogecoin case) contributes to the discussion, it plays a more marginal role compared to the data-driven Reddit study, which constitutes the main focus of this thesis. The methodology used for these analyses will be described in more detail in the dedicated section.

The thesis is divided into three main chapters: the first one introduces the topic of the paper, provides a general overview of behavioural finance and informs the reader about the rise of social networks from a financial perspective; the second one analyses various literatures by reviewing articles on topics fundamental to the study being conducted such as behavioural finance or the relationship between social media and the financial market and other topics aligned with the aim of the thesis. Lastly, the final chapter contains the qualitative and quantitative analysis discussed above, leading to the conclusions of the study.

1.2 Behavioural finance overview and investor decision-making

For the most part of the 20th century, mainstream financial theory rested upon the assumptions that investors act rationally, and markets are efficient. This idea is notoriously related to the work of Eugene Fama, due to his 1970 review of theoretical and empirical research[1]. The view that he

[1] Fama, E. F. (1970). *Efficient Capital Markets: A Review of Theory and Empirical Work*. Journal of Finance, 25(2), 383–417.

formalised is the so-called **Efficient Market Hypothesis (EMH),** which states that security prices always "fully reflect" available information. He even presents a differentiation, firstly suggested by Harry Roberts, between *weak form* efficient markets, where the information set is just historical prices, *semi-strong form*, in which the information set also contains the publicly available information, and the *strong form* where all relevant information for price formation purposes is considered. In this model, investors make unbiased forecasts and select optimal portfolios, following the modern **Capital Asset Pricing Model (CAPM)**[2] hypothesis. In general, the standard finance foundation is based on the contributions of, other than the above-mentioned Fama, mainly Merton Miller and Franco Modigliani, Harry Markowitz, William Sharpe…

However, several empirical anomalies began to challenge the universality of these assumptions. High volatility, sudden market crashes, and speculative bubbles seemed to contradict those fundamental theories. An example of this is notable in Shiller's work[3], which openly critiques the notorious Fama's 1970 article by writing that *"it did report some anomalies like slight serial dependencies in stock market returns."* Moreover, he deeply analyses the excess volatility in the 1980s markets and shows that price swings were far greater than what the fundamental theories could justify.

These kinds of inconsistencies pushed scholars to explore alternative explanations, and so they began to acknowledge the impact of psychological factors on decision-making. This shift marked the birth of **behavioural finance**, a field that merges finance with cognitive psychology and emotions. In this context, behavioural finance emerged not as a radical rejection of traditional theory, but as a necessary evolution.

The first research in this field could be attributed to Adam Smith, the author of the famous "invisible hand" theory. In his first book[4], he discusses how people's behaviour is guided not only by passions, empathy (which he calls "sympathy") and the desire for social approval ("*Man naturally desires, not only to be loved, but to be lovely"*). The work is packed with ideas related to what is referred to as loss aversion, intertemporal choice, and overconfidence[5]. He even appears aware of concepts like cognitive dissonance, status-seeking and irrational actions, all of which are foundations of today's behavioural economics.

---

[2] Sharpe, W.F. (1964) *Capital Asset Prices: a Theory of Market Equilibrium Under Conditions of Risk*. Journal of Finance 19(3), 425-442.
[3] Shiller, R. J. (2003). *From Efficient Markets Theory to Behavioral Finance*. Journal of Economic Perspectives, 17(1), 83–104.
[4] Smith, A. (1759). *The Theory of Moral Sentiments*.
[5] Ashraf, N., Camerer, C. F., & Loewenstein, G. (2005). *Adam Smith, Behavioral Economist*. Journal of Economic Perspectives, 19(3), 131–145.

Shifting to the 20th century, the work of George Selden[6] must be mentioned, as it directly explored investor behaviour and how it's driven by emotions and psychological factors. He firmly believed that "*the movements of prices on the exchanges are dependent to a very large degree on the mental attitude of the investing and trading public.*"

The next big discovery in the behavioural finance field is attributed to Herbert Simon, who led the transition from a classic economic rationality to a more modern depiction of human decision-making. He theorised a **bounded rationality**[7], arguing that humans make decisions under cognitive constraints such as limited information, time or mental processing capacity: "*…the task is to replace the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed…"*. His critiques of the neoclassical rational agent culminated in his 1978 Nobel Prize in Economics[8] where he was recognised for *"for his pioneering research into the decision-making process within economic organizations"*. He emphasises again his bounded rationality theory by reporting the real-life limitations to human decision-making. A key extension of this idea, which was reported in his lecture, came from his studies with Cyert and March[9] where organizations were portrayed not as fully rational entities but as coalitions of individuals with conflicting goals, limited information, and satisficing behaviour.

Simon's ideas turned out to be beneficial for laying the groundwork for the discoveries of Daniel Kahneman and Amos Tversky[10],. These two psychologists would go on to significantly deepen the understanding of decision-making under uncertainty, continuing Simon's work and directly challenging the rational assumptions embedded in traditional finance. Their collaboration became a cornerstone of behavioural finance, as they explored how cognitive limitations shape human judgment in predictable, yet systematic ways. In their notorious 1974 paper they identified a set of intuitive mental shortcuts, also known as **heuristics**, that humans rely on when facing complex decisions. Some examples might be availability heuristic, which consists in evaluating the likelihood of events based on how easy examples come to mind or relying too heavily to an "anchor" (an initial piece of information) when making decisions. They discovered that those heuristics weren't merely some random errors, but that were plausible behaviours driven by

6 Selden, G. C. (1912). *Psychology of the Stock Market: Human Impulses Lead to Speculative Disasters*. New York: Ticker Publishing.
7 Simon, H. A. (1955). *A Behavioral Model of Rational Choice*. The Quarterly Journal of Economics, 69(1), 99–118.
8 Simon, H. A. (1978). *Rational Decision-Making in Business Organizations*. Nobel Prize Lecture.
9 Cyert, R. M., & March, J. G. (1963). *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall.
10 Kahneman, D., & Tversky, A. (1974). *Judgment under Uncertainty: Heuristics and Biases*. Science, 185(4157), 1124–1131.

structural informational limits and psychological constraints. Their most profound theoretical contribution came in 1979, with their Prospect Theory[11], which proposed that people evaluate outcomes relative to a reference point and exhibit **loss aversion**. presented as an alternative to the Expected Utility Theory, and it tries to explain numerous real-world phenomena like panic-selling or reluctance to cut losses. A more detailed overview of these psychological aspects is discussed in Chapter 2.1. Their findings reshaped the foundations of economics and finances so much, that in 2002, Daniel Kahneman was awarded the Nobel Prize in Economic Sciences *"for having integrated insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty"*.

In summary, the transition from classical financial models to behavioural finance has redefined our understanding of investor decision-making. These theoretical insights illuminate why investors often make suboptimal decisions and they establish a complex framework that not only questions the assumptions of market efficiency but also provides a more realistic picture of investor behaviour.

This robust theoretical framework sets the stage for the subsequent empirical analyses. In the following chapters, the focus will shift to exploring specific behavioural mechanisms, in accordance with the thesis goal to connect these foundational theories with observable market outcomes, ultimately enriching our understanding of the interplay between online sentiment and financial market dynamics.

1.3 The rise of social network in Finance

Until the 90s, the access to information was significantly limited, highly centralised and costly. The main sources were:

- traditional financial media (The Wall Street Journal, Financial Times, CNBC)
- analyst reports
- academic and technical publications
- official corporate channels.

Given these premises, it can be assumed that all the information came from professional investors, limiting small investors to a more passive approach, and so it created a structural information asymmetry. In fact, most of these institutional actors were inaccessible to the average investors

[11] Kahneman, D., & Tversky, A. (1979). *Prospect Theory: An Analysis of Decision under Risk*. Econometrica, 47(2), 263–291.

due to their costs. These financial media played a key role in shaping market sentiment and guiding investor behaviour.

One of the most influential studies in this area is by Paul Tetlock[12] , who demonstrates that the tone of daily columns in The Wall Street Journal significantly affects market returns and trading volumes.

A huge switch to this perspective occurred in the late 90s with the rise of forums and message boards such as Yahoo! Finance, RagingBull, The Motley Fool... For the first time, investors gained platforms where they could share opinions, speculate on markets, and crowdsource insights without relying on institutional sources or traditional financial media. These forums represented a horizontal form of communication, where knowledge could flow peer-to-peer, challenging the traditional top-down model dominated by analysts, brokers, and financial journalists. Despite their informal appearance and occasional unreliability, these spaces made investors feel more empowered.

A landmark study by Antweiler and Frank[13] analysed over 1.5 million messages posted on Yahoo! Finance and RagingBull and found that message board activity significantly correlated with both stock trading volume and volatility. The study clearly concluded that "*Message posting activity is a readily observed correlate that helps to account for volatility*." It also suggested that online chatter was not mere noise but a driver of market behaviour.

The 2010s marked a pivotal transition from static forums to dynamic social media platforms, radically transforming the way financial information is created, shared, and consumed. Unlike earlier forums, platforms like Twitter, Reddit, YouTube, and later Instagram and TikTok, enabled users not only to post and discuss but also to build a following, develop personal brands, and broadcast opinions in real time, dramatically increasing the reach and speed of financial content.

Twitter is a microblogging service favoured by both retail and professional investors, fostering fast-paced finance-related conversations. Already in 2014 Sprenger et al.[14] found that researchers and financial professionals could use tweet features as valuable proxies for investor behaviour and

---

[12] Tetlock, P. C. (2007). *Giving Content to Investor Sentiment: The Role of Media in the Stock Market*. The Journal of Finance, 62(3), 1139–1168

[13] Antweiler, W., & Frank, M. Z. (2004). *Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards.* The Journal of Finance, 59(3), 1259–1294

[14] Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). *Tweets and Trades: The Information Content of Stock Microblogs.* European Financial Management, 20(5), 926–957

belief formation. Influential voices, ranging from hedge fund managers to independent traders, began to emerge, building credibility and sometimes even moving markets with a single tweet.

Meanwhile, Reddit introduced a different model of community-driven financial engagement. This social network can be considered as a collection of communities, so-called subreddits, each one with its own culture and rules. For the financial lovers, subreddits like r/investing, r/stock or r/wallstreetbets quickly became popular. More about Reddit specifically will be discussed in the next chapters.

At the same time, YouTube became the home of finfluencers, long-form content creators, who discussed stocks, reviewed investing concepts and shared their trading strategies. In recent years, TikTok and Instagram would amplify this trend of short-form viral content that made financial discourse more digestible (at the cost of reduced accuracy).

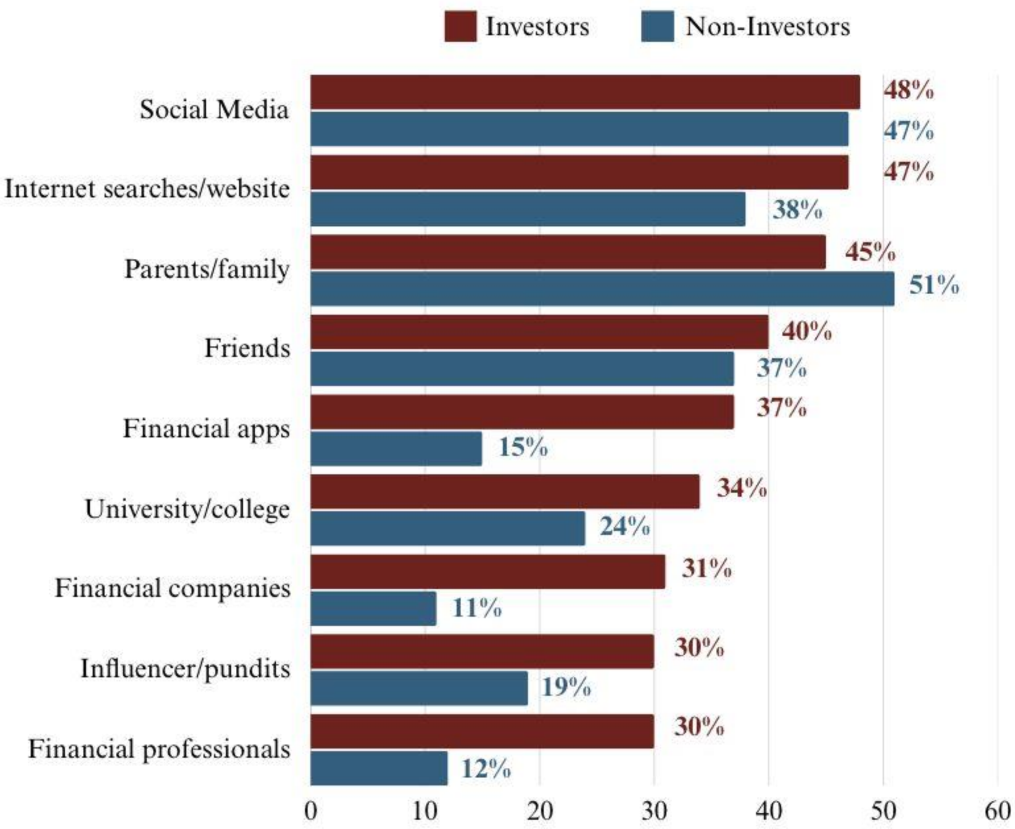
This new figure, the finfluencer, embodies a hybrid role between educator, entertainer, and financial commentator. Their popularity surged especially after the 2008 financial crisis and during the COVID-19 pandemic, when traditional financial trust collapsed. This led retail investors to search for relatable, accessible sources of financial knowledge. As already emphasised before, finfluencers often simplify complex concepts like options trading or portfolio diversification into appealing narratives or visually stimulating videos, creating a new channel of financial literacy.

Unlike traditional financial advisors, most finfluencers operate without formal certification or with some light regulatory oversight, democratising access to finance. This has drawn in millions of younger investors: platforms like TikTok have proven particularly influential; hashtags such as #stocktok or #stocktips accumulated billions of views, showing the power of viral trends in shaping investor decisions, while also raising several concerns over their reliability.

A 2023 study[15] by the CFA Institute and the FINRA Investor Education Foundation found that 48% of U.S. Gen Z investors use social media to learn about investing, making it the top information source for this demographic. This trend underscores the shifting landscape of financial education, where traditional barriers are lowered and information is democratised.

[15] CFA Institute & FINRA Investor Education Foundation. (2023). *Gen Z and Investing: Social Media, Crypto, FOMO, and Family.*

Figure 1 - Sources of Information Gen Zs Use to Learn (source:CFA)



The results showed in Figure 1 represent a significant proof of the psychological power that finfluencers and, more generally, social media could have in shaping investor behaviour, sometimes leading to impulsive financial decisions. The same study also reports the factors that motivate the same U.S. GenZs to finally invest:

Table 1 - Factors that influenced the decision to invest (source:CFA)

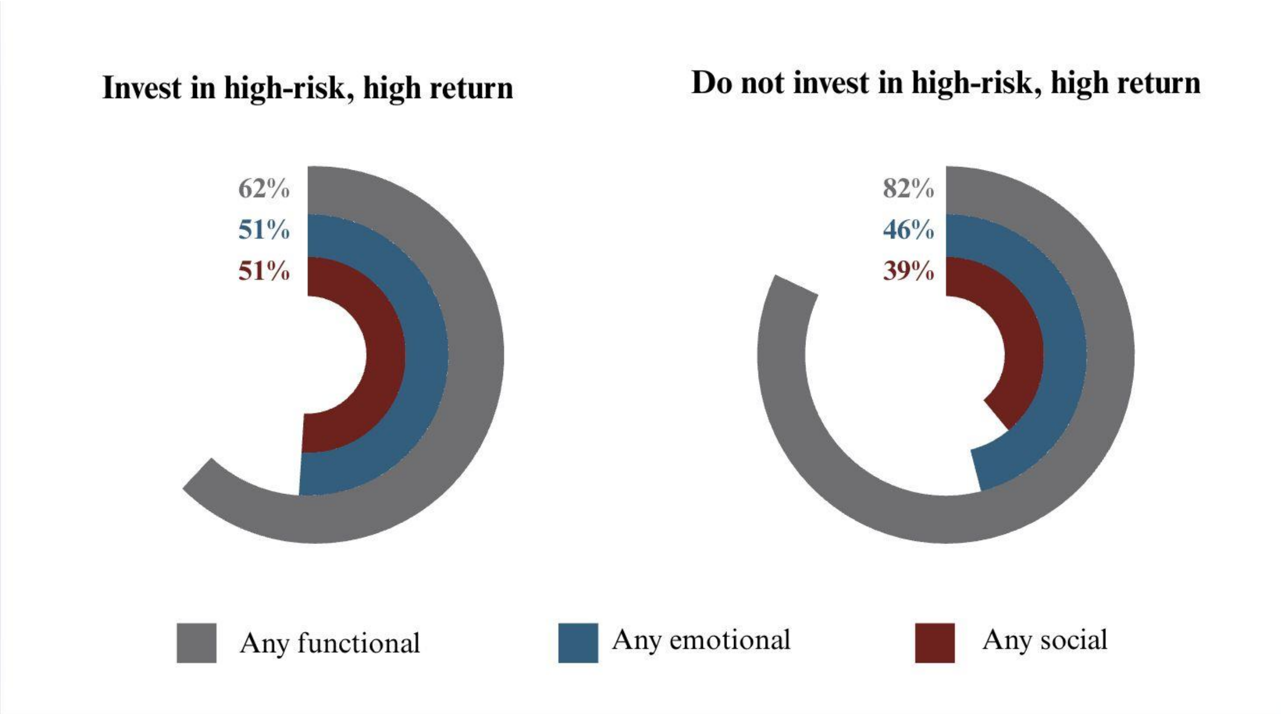
Major Factor in Decision to Invest (4 or 5 on a 5-point scale)	U.S. Gen Z Investor
Ability to start investing with small amounts	67%
Curiosity/own interest	65%
Obtained money to invest	57%
Parent/family member	54%
Ease of opening account	53%
Promotional incentive (cash, crypto, stock)	45%
Friends/colleagues	44%
Fear of missing out (FOMO) on opportunity to grow money	41%
Social media influencer/pundit	37%
Teacher/professor	31%
Online discussion board	31%
Advertisement/push notification	30%
My employer	29%

Not surprisingly, as showed in Table 1, almost 41% of the investors cited fear of missing out as a factor in their decision to start investing, a clear suggestion of the above-mentioned impact of finfluencers. Even the most adamant would agree that this category would need an in-depth regulatory framework. Despite these drawbacks, finfluencers have undoubtedly altered the landscape of modern finance, often providing valuable entry points for financial literacy, showing how behavioural and social dynamics are shaping the way investors think and act.

This rapid evolution in financial communication has coincided with a profound transformation in the profile and motivations of the typical retail investor. The archetype of the prudent, long-term investor guided by traditional institutional advice is being complemented, by a younger, more autonomous, and fully digitally immersed generation. These investors often view markets through a lens of community, identity, and even entertainment, rather than solely as vehicles for wealth preservation, retirement planning, or hedging purposes. The aspect of gamification is, in fact, quite a peculiar one to the younger generations. They're attracted by game-like mechanics, like a ranking system or a task/reward connection, and this massively increases their engagement with financial platforms. For example, Robinhood and eToro are known for their slick interfaces and game-like features and so contribute to this transformation by lowering the barriers to market participation and creating a sense of agency and control among retail traders.

According to the UK Financial Conduct Authority[16], a great number of new retail investors are motivated not just by rational financial goals, but by emotional and social factors, such as the thrill of investing, the status it may confer, or the sense of connection and competition with online communities.

Figure 2 - Top three motivators for investing by self-directed investor risk type (source:FCA)



It immediately catches the viewers' eyes how much more important social and emotional motivators are when it comes to investing rather than not investing. This behavioural pivot can easily be related to the influencer arguments previously proposed.

This newer investor archetype would not appear inherently less competent but is often more reactive, socially engaged, and psychologically exposed to information overload and unconscious behavioural triggers. The financial behaviour may well reflect the theories proposed in the previous paragraph.

To sum up, the modern investor is no longer a static profile but a composition of digitally empowered, emotionally driven individuals whose decision-making processes reflect the complex interplay of psychology, technology, and culture. This shift represents a major upheaval to the assumptions of traditional finance, as has been previously observed, prompting regulators to reconsider what it means to invest in the digital age.

[16] FCA (2021). *Understanding self-directed investors: Research and findings.* Financial Conduct Authority.

2.LITERATURE REVIEW

2.1 Behavioural finance and the role of emotions in investments (cognitive, biases herding...)

In paragraph 1.2, the dissonances between classical financial theories and real-world investor behaviour were introduced. To assess the validity of the behavioural critique, it would be necessary to determine whether asset prices accurately reflect the information available. This, in turn, requires acknowledging the difficulty of devising measures of a security's intrinsic value due to a series of complex analytical steps, such as forecasting future cash flows, choosing a fair discount rate, intangible firm-specific assets[17]…

Therefore, the literature chose to focus on two kinds of tests[18]: the anomalies literature that examined the potential risk-adjusted returns from strategies like investing in stocks with momentum or in value, or the results of actual investments, and so comparing the track records of managers and comparing them to the market performance. Neither class of tests has proven fully conclusive.

Building upon the EMH critiques, a more granular exploration of the psychological forces that drive financial decisions is now undertaken. This paragraph aims to build a comprehensive framework of investor irrationality by dissecting processing errors, behavioural biases and revising limits to arbitrage. In fact, the existence of limits to arbitrage is pivotal to the behavioural finance theories; otherwise, the lone presence of irrationalities that affect prices would be rapidly taken advantage of by sharp-eyed arbitrageurs. Reasoning on these points, it is immediately apparent how these premises lead to a vibrant conclusion. Indeed, while it is true that if the price of an asset is equal to its value, then the market is efficient, it cannot be said that if there are no systematic profit opportunities then the market is efficient, precisely because of the limitations of arbitrage mentioned above.[19]

Having established the theoretical space for inefficiencies, the focus now shifts to the errors that might occur in **information processing**. The core of many market anomalies often lies in the decisions that investors make when internalising information in the first place. In this process, in fact, people are extremely vulnerable to cognitive constraints and dissonances.

[17] Lev, B., & Gu, F. (2016). *The End of Accounting and the Path Forward for Investors and Managers* (1st ed., Chapter 8). Wiley.
[18] Bodie, Z., Kane, A., & Marcus, A. J. (2021). *Essentials of Investments* (12th ed., Chapter 9). McGraw-Hill Education.
[19] Barberis, N., & Thaler, R. (2003). *A Survey of Behavioral Finance*. Handbook of the Economics of Finance: Volume 1B – Financial Markets and Asset Pricing, 1053–1128.

To start our information processing error review, it must be stated that people have **limited attention**: since our cognitive bandwidth is finite, individuals tend to rely on rules of thumb and heuristics. A well-known example of this is reported by Sloan\[20] with his accrual anomaly: he found that firms that exhibit a high accrual component in their earnings usually experience lower future stock returns. Conversely, firms with lower accruals (where reported earnings more closely reflect cash flows) tend to outperform. These findings suggest that investors don't accurately scrutinise the financial statements deeply enough to actually understand the real worth of a company. Evidently, limited attention causes **underreaction** and **overreaction**. Indeed, the investors that really don't have quality information, as in the accruals case, might not react appropriately by misinterpreting salient signals and thus underreacting. By contrast, recent dramatic events, like market crashes or positive earnings news, might provoke excessive movements. Important confirmation of this phenomenon is retrievable in the works of De Bondt and Thaler (1985[21] and 1987[22]). In their first study, they divided several stocks into two categories, "winners" and "losers", based on their past 3 years performance. The results were shocking: *"Thirty-six months after portfolio formation, the losing stocks have earned about 25% more than the winners, even though the latter are significantly more risky"*. This example firmly supports the **overreaction** error by demonstrating how investors potentially overweigh and overestimate positive financial signals.

In continuity with overreaction, another important information processing error is **overconfidence**, a tendency for individuals to overestimate the accuracy of their information, forecasts, or abilities. First, people overrate their ability to estimate quantities[23] and, second, they poorly calibrate probabilities. For example, they categorise events that happen 80% of the time as certain and the ones that happen 20% of the times as impossible[24]. It is possible that the overconfidence bias derives from other two biases: the hindsight and self-attribution ones. The first refers to the people's tendency to believe, after an event has occurred, that they successfully predicted it beforehand. This could easily lead to erroneous conclusions, resulting in a false confidence in their predicting skills, often making them think they could even predict the future more accurately. The

---

[20] Sloan, R. G. (1996). *Do stock prices fully reflect information in accruals and cash flows about future earnings?* The Accounting Review, 71(3), 289–315.
21 De Bondt, W. F. M., & Thaler, R. (1985**).** *Does the stock market overreact?* The Journal of Finance, 40(3), 793–805.
22 De Bondt, W. F. M., & Thaler, R. (1987*). Further evidence on investor overreaction and stock market seasonality.* The Journal of Finance, 42(3), 557–581.
[23] Alpert, M., & Raiffa, H. (1982). *A progress report on the training of probability assessors*. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under Uncertainty: Heuristics and Biases (pp. 294–305). Cambridge: Cambridge University Press.
[24] Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). *Knowing with certainty: The appropriateness of extreme confidence*. Journal of Experimental Psychology: Human Perception and Performance, 3(4), 552–564.

self-attribution bias refers to the inclination to take full personal recognition of any success and to blame the failures on uncontrollable circumstances (Barberis & Thaler, 2003). A particularly revealing study by Barber and Odean (2001)[25] provides other empirical evidence of overconfidence in retail trading behaviour. Analysing brokerage accounts, they found that especially single men traded significantly more frequently than women. This aligns with broader psychological findings showing that men exhibit higher levels of overconfidence and selfishness. Crucially, Barber and Odean also discovered that higher trading frequency was inversely correlated with portfolio: the top 20% of accounts ranked by portfolio turnover had average returns seven percentage points lower than the 20% of the accounts with the lowest turnover rates. Their study famously concluded: *"Trading is hazardous to your wealth."*. Even in M&A contexts, signs of overconfidence can be found overconfident CEOs are more likely to overpay for target firms[26].

Similar to overconfidence is **optimism**, in fact most people have unrealistic views of their abilities and about their future life events[27]. A famous survey states that 90% of drivers in Sweden ranked themselves as better-than-average drivers and similar data is detectable in domains such as the ability to get along with people and sense of humour.

A **conservative bias** occurs when investors are too attached to old beliefs and they don't update rapidly to new evidence, so people might initially underreact to news about a firm, so that prices will fully reflect new information only gradually. Such a bias would give rise to momentum in stock market returns. A study by Edwards[28] (1968) analysing the individuals' estimate accuracy, proves that usually people react too little when data is not representative of a particular model and too much if the data is backed by an underlying model. Using Bayesian updating tasks in experimental settings, he found that individuals significantly underweighted new evidence, holding onto prior beliefs more strongly than a rational Bayesian model would predict.

Another crucial error consists of the extrapolation of illusionary patterns, and it's called **representative** bias. Kahneman and Tversky show this fallacy with the notorious Linda problem[29]. The example goes:

---

[25] Barber, B. M., & Odean, T. (2001). *Boys will be boys: Gender, overconfidence, and common stock investment*. The Quarterly Journal of Economics, 116(1), 261–292.
[26] Malmendier, U., & Tate, G. (2008). *Who makes acquisitions? CEO overconfidence and the market's reaction*. Journal of Financial Economics, 89(1), 20–43.
[27] Weinstein, N. D. (1980). *Unrealistic optimism about future life events*. Journal of Personality and Social Psychology, 39(5), 806–820.
[28] Edwards, W. (1968). *Conservatism in human information processing*. In Kleinmutz, B. (Ed.), Formal Representation of Human Judgment . New York: Wiley, 17-52.
[29] Tversky, A., & Kahneman, D. (1981). *Judgments of and by representativeness* (Report). Stanford University.

"Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and also participated in anti-nuclear demonstrations.

Which is more probable?

1. *Linda is a bank teller.*
2. *Linda is a bank teller and is active in the feminist movement."*

When asked this question, individuals often answer the latter, because the second option seems to better fit the description of Linda. Another consequence of the representative bias is the tendency of people not to consider the size of samples. When subjects don't already have information on something, they will, at least in the first instance, infer some kind of beliefs, not caring about the sample size of the data. An example of that is the "hot hand" phenomenon in basketball[30]: when a player hits a few shots in a row, people will overestimate the probability of his next shot to go in. In finance this error could impact investors' expectations, and this may lead to a gap between prices and intrinsic values.

The **anchoring** error happens when people, while estimating values, rely too heavily on an initial arbitrary value, and then they adjust their estimate away from it. In a notorious experiment, the participants were asked to estimate the percentage of African countries in the United Nations. Before providing their estimates, they were asked whether they think their guess was higher or lower than a randomly generated number. Surprisingly, the successive estimates were heavily influenced by the arbitrary number (Kahneman and Tversky, 1974). More recently, Ariely, Loewenstein, and Prelec (2003)[31] conducted a study where participants were first asked to write down the last two digits of their social security number. Then, they were shown various products and asked whether they would be willing to pay the amount of their two-digit number for each item. After that, they were asked how much they would actually pay. People with higher social security digits ended up stating they would pay significantly more for the items than those with lower digits. Even though the social security number is completely random and irrelevant to the product's value, it acted as an anchor that influenced their willingness to pay.

[30] Gilovich, T., Vallone, R., & Tversky, A. (1985*). The hot hand in basketball: On the misperception of random sequences.* Cognitive Psychology, 17(3), 295–314.
[31] Ariely, D., Loewenstein, G., & Prelec, D. (2003). *Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences.* The Quarterly Journal of Economics, 118(1), 73–105.

Even if people manage to perfectly acquire information without errors, they might still be subject to **behavioural biases**. They are psychological tendencies that influence investor behaviour, and they often originate from deeper motivational or emotional forces.

People's decisions might be often subject to a **framing** bias. This means that their choice might differ based on how the question is posed. An individual might reject a bet if posed evidencing the risk-reward ratio, but he may accept if it's framed evidencing the winning amounts. A question-framing example originates from Howard Rugg, a sociologist in the 1940s, who illustrated how minor wording changes in survey questions can significantly alter respondents' answers. He showed how 62% of people disagreed with allowing "public condemnation of democracy", but only 46% of people agreed that it was right to "forbid public condemnation of democracy"[32],

A specific form of framing is **mental accounting**: people tend to treat money differently depending on the origin or the intended use, not considering that money is fungible. The bias was initially formalised by Richard Thaler (1980). A fascinating experiment proposed by Kahneman and Tversky (1984) beautifully explain the concept:

"Problem 8 (N = 200): Imagine that you have decided to see a play and paid the admission price of $10 per ticket. As you enter the theater, you discover that you have lost the ticket. The seat was not marked, and the ticket cannot be recovered.

Would you pay $10 for another ticket?

Yes (46%) No (54%)

Problem 9 (N = 183): Imagine that you have decided to see a play where admission is $10 per ticket. As you enter the theater, you discover that you have lost a $10 bill.

Would you still pay $10 for a ticket for the play?

Yes (88%) No (12%)[33]*"*

The cause of this absurd results probably come from the logic that individuals follow in their buying process. Since they view the activity of going to the theatre as a transaction, they might think that destining $20 to that specific experience is too much, while they would more likely buy the ticket in problem 9, because they feel that the loss of the bill is unrelated to the activity.

[32] Plous, S. (1993). *The Psychology of Judgment and Decision Making*. New York: McGraw-Hill.
[33] Kahneman, D., & Tversky, A. (1984). *Choices, values, and frames*. American Psychologist, 39(4), 341–350.

Closely related to mental accounting is the **disposition effect**. It can be defined as the tendency of investors to sell winning investments too early and hold onto losing investments too long. A behavioural explanation is that an average investor might mentally "open" an account for each stock and experience the pain of realising a loss more strongly than the pleasure of realising a gain, leading to irrational holding behaviour. A classic study by Shefrin and Statman (1985)[34] formalised this effect, and they showed how gains and losses are evaluated relative to a reference point (typically the purchase price), and investors are more likely to realise gains because they secure a psychological profit, while holding losers avoids the pain of loss realisation. This not only contradicts standard economics theory, but also directly contrast a tax-minimisation strategy.

Strongly associated is the **house-money effect**, which describes how investors become more willing to take risks after experiencing prior gains. In an utterly gambling behaviour, people tend to treat their profits as "house money" and therefore are less loss-averse with it, leading to an increasing likelihood of risk-seeking behaviour in following decisions[35].

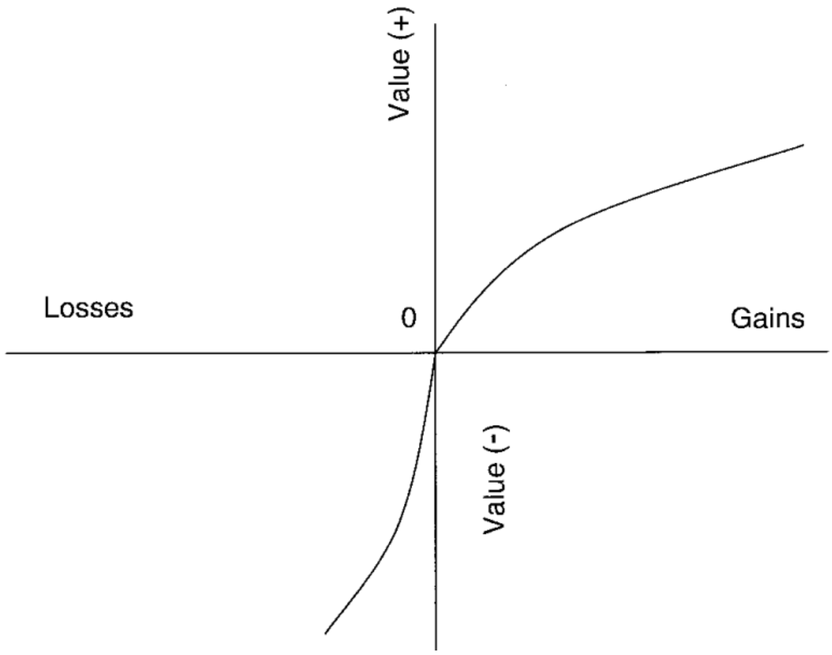
One of the most representative formalisations of these biases is the one proposed by Kahneman and Tversky in 1979, already introduced in paragraph 1.2: **prospect theory**. Their study convincingly shows that people evaluate choices relative to a reference point, deviating from expected utility theory. **Loss aversion** is one of the main pillars of the model; it is indicated that individuals experience the pain of losses more intensely than the pleasure of equivalent gains, similarly to what was already discussed for the disposition effect. This is visually represented by an S-shaped value function (Figure 3), concave for gains and convex for losses, with a steeper slope for losses, indicating that a €100 loss feels worse than a €100 gain feels good.

---

[34] 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.

[35] Thaler, R. H., & Johnson, E. J. (1990). *Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice*. Management Science, 36(6), 643–660.

Figure 3 - Kahneman and Tversky value function (source: Kahneman and Tversky (1975))



Another crucial component of the prospect theory is **probability weighting**: individuals overweigh small probabilities (which helps explain the appeal of lotteries or insurance) and underweigh high probabilities, leading to decisions that deviate systematically from rational benchmarks. To better explain the concept, reference will now be made to Problem 3 and 4 illustrated in the above-mentioned study.

Problem 3:

Option A: A sure gain of $3000

Option B: An 80% chance to win $4000, and a 20% chance to win nothing

Problem 4:

Option C: A sure loss of $3000

Option D: An 80% chance to lose $4000, and a 20% chance to lose nothing

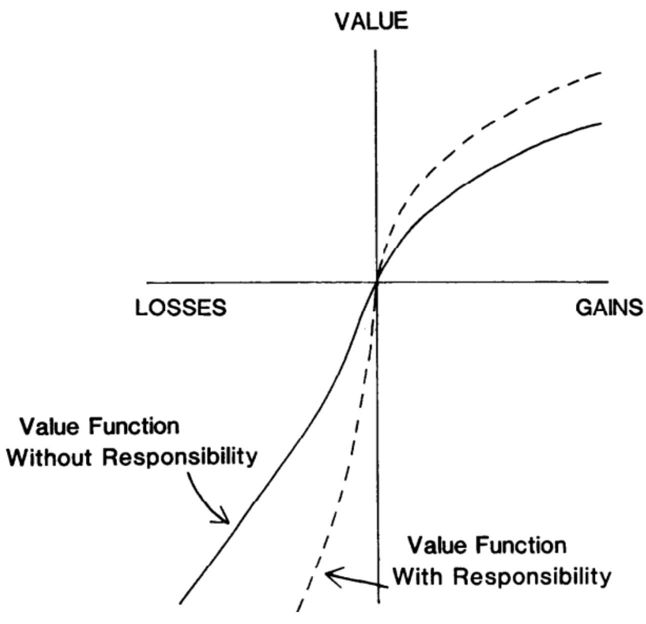
In Problem 3 even though Option B has a higher expected value ($4000 × 0.8 = $3200), people underweigh the high probability of winning and overvalue the certainty of Option A. This is obviously correlated to risk aversion, driven not just by the shape of the value function, but by the fact that the 80% chance feels less than 80% psychologically. In Problem 4, instead, the expected loss in Option D is **$3200**, which is worse than the sure loss of $3000. Yet, people still choose the risky option, overweighing the small probability of losing nothing, making Option D seem more attractive. This type of risk-seeking is the same one shown in the previously mentioned disposition effect.

Regret avoidance refers to the investors preference to make decisions in a way that minimises the potential for future regret, rather than maximising utility or expected returns. Kahneman and Tversky (1982)[36]explain this bias with an example:

"Paul owns shares in Company A. During the past year he considered switching to stock in Company B, but he decided against it. He now finds that he would have been better off by $ 1 ,200 if he had switched to the stock of Company B. George owned shares in Company B. During the past year he switched to stock in Company A. He now finds that he would have been better off by $ 1 ,200 if he had kept his stock in Company B. Who feels more regret? H ere again it is generally agreed that George is more upset than Paul, although their objective situations are now identical (both own the stock of Company A) and each reached his situation by deliberate decision.

Apparently it is easier for George to imagine not taking an action (and there by retaining the more advantageous stock) than it would be for Paul to imagine taking the action. Furthermore, one would expect both men to anticipate the possibility of regret and to act accordingly. In general the anticipation of regret is likely to favor inaction over action and routine behavior over innovative behavior."

Figure 4 - Individual value function (source: Shefrin, H., & Statman, M. (1985))



[36] Kahneman, D., & Tversky, A. (1982). *The Psychology of Preferences*. Scientific American, 246(1), 160–173.

Shefrin and Stateman (1985) provide a graph of the individual value function implementing the 1979 Kahneman and Tversky one. It's evident in the graph above that the regret effect (measured by the vertical distance between the two functions in the third quadrant) is stronger than the pride effect (measured by the vertical distance between the two functions in the first quadrant). This perfectly explain why people psychologically tend to avoid regrets.

These examples represent only a partition of the numerous behavioural biases and information seeking errors studied and identified by researchers, but it's a sufficient start to make the reader aware of the foundational ideas and principles of behavioural finance theory.

Beyond individual-level biases, financial behaviour is also shaped by collective dynamics. Among the most pervasive is **herding**. The herding effect in behavioural finance represents a foundational concept tracing back to early economic thought, with its formalisation emerging through interdisciplinary research bridging psychology, sociology, and economics. The phenomenon, where individuals mimic collective actions rather than relying on independent analysis, was first intended by Keynes with is famous beauty contest[37] example:

"...professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view. It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth and higher degrees."

This metaphor brilliantly underlines how investors prioritise anticipating peers' actions over intrinsic asset value, fostering self-reinforcing conventions detached from fundamentals. Herding behaviour was later systematically theorized in the 1990s. Scharfstein and Stein[38] (1990) introduced reputational herding, arguing that fund managers might imitate peers to prioritise career security over accurate decision-making. and also because of psychological factors like the

[37] Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money*. London: Macmillan, Ch. 12.
[38] Scharfstein, D. S., & Stein, J. C. (1990). *Herd behavior and investment*. The American Economic Review, 80(3), 465–479.

"sharing-the-blame" effect- as the researchers call it. Later, a 1992 paper[39], laid the groundwork by theorising informational cascades, where agents successively abandon private information to follow predecessors' decisions, leading to suboptimal market outcomes. This model demonstrated that herding could arise rationally in environments with asymmetric information, as individuals infer hidden signals from others' actions, even if those actions are based on incomplete data. The year later Kirman[40] showed how an agent-based model of ant recruitment illustrated herding as a stochastic process, analogous to traders following market trends without fundamental justification.

While these biases explain why mispricings may emerge, they do not by themselves explain why these inefficiencies persist in a competitive market. From a theoretical standpoint, as introduced before, behavioural finance requires that arbitrage opportunities be limited or risky, otherwise rational traders would exploit those wrong prices until they vanish. Without such limits, the presence of irrational traders would have no lasting effect on prices, as markets would self-correct. This is why the concept of limits to arbitrage is fundamental: it provides the necessary condition for behavioural anomalies to have real and persistent market impact. In an important 1997 paper[41] Shleifer and Vishny concluded:

"More generally, specialized, professional arbitrageurs may avoid extremely volatile "arbitrage" positions. Although such positions offer attractive average returns, the volatility also exposes arbitrageurs to risk of losses and the need to liquidate the portfolio under pressure from the investors in the fund."

This directly leads us to **fundamental risk:** even if a security is underpriced, there is still a risk related to buying it, because there's no guarantee that the price will meet the fair value in the buyer's investment horizon. A classic example is the Royal Dutch-Shell one. In 2005, Royal Dutch Petroleum and Shell Transport & Trading were still operating as two separately listed companies despite being part of the same corporate entity. They had agreed to share profits at a fixed 60:40 ratio. However, arbitrageurs noticed that Royal Dutch stock was often overvalued relative to Shell and attempted to exploit the spread by shorting Royal Dutch and buying Shell. Even though, this seemed like a clear arbitrage opportunity, the spread persisted for years, and in some cases even widened, leading to losses for arbitrageurs despite being theoretically correct in their bet. That happened because stock prices could be determined by unpredictable factors and moreover, there was no certainty as to when the prices would converge, so investors who entered early could suffer

[39] 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.
[40] Kirman, A. (1993). *Ants, rationality, and recruitment*. The Quarterly Journal of Economics, 108(1), 137–156.
[41] Shleifer, A., & Vishny, R. W. (1997). *The Limits of Arbitrage*. *The Journal of Finance*, 52(1), 35–55.

interim losses or margin calls, especially if they were managing external capital and had to report short-term performance.

Other common limitations are the so-called **implementation costs**. These costs include transaction fees, short-selling constraints and borrowing fees, taxes or capital requirements… In general, they consist of real-world frictions that make exploiting mispricing costly and, at times, unprofitable.

Model risk refers to the possibility that the arbitrageur's valuation model might be incorrect, so, more generally, that the market is right and the arbitrageur is wrong. Even when a trader believes an asset is mispriced, the mispricing may reflect real, unknown information that their model fails to account for. This is a subtle but powerful source of risk, because it challenges the very premise of arbitrage, but also because the misjudgement may remain hidden for a long time, potentially causing massive losses. An example is the valuation of cryptocurrencies like Bitcoin or Ethereum, because they don't produce cashflows or dividends, and so valuing them is a hard task. Since multiple valuation models have been used and there has been a certain degree of unpredictability in the evaluations, it's safe to assume that in crypto the model risk is quite high.

Taken together, the cognitive errors, behavioural biases, and structural barriers to arbitrage examined in this chapter demonstrate that markets are far from the frictionless and rational ecosystems proposed by classical theory. These insights will serve as a conceptual bridge to the next chapters, where the exploration focuses on how modern digital platforms and social media environments intensify behavioural patterns.

2.2 Relations between social media and financial markets

The interdependent relationship between social media and financial markets has redesigned modern investment ecosystems, creating unique channels for information dissemination and collective action. The paragraph synthesises empirical evidence, behavioural theories, and regulatory challenges arising from digital platforms' influence on market dynamics and it will also analyse pivotal events like the GameStop short squeeze.

The dynamics of financial information streams endured radical transformation with social media's ascendance, dismantling institutional gatekeepers that traditionally filtered market narratives. This democratisation has enabled retail investors to participate in real-time financial discussion through platforms optimised for emotional engagement rather than rational technical analysis. To cite Robert Shiller's works, it's fascinating to refer to his concept of "**narrative economics**": in his study[42] he highlights how viral narrations function much like epidemics, spreading not because of their adherence to the truth, but because they are emotionally resonant and easily disseminated. The repetition of simplified, emotionally charged narratives on social media mirrors what the American economist identifies as a key trait of effective economic storytelling: their ability to convey identity, interest, and a sense of urgency. Indeed, the distinguishing features that magnify social networks' market impact are temporal compression, emotional amplification, and social validation.

Reconnecting to behavioural finance theory, reference can again be made to Simon, who, in his bounded rationality framework, specifically affirms:

"In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it."[43]

It's evident the attention problem in the social media era is even greater than it was in the Simon's times, so his lesson can be reinterpreted in a modern context and be synthesised as "**content abundance**, **attention scarcity**". Contemporary research shows that this cognitive burden effectively affects financial markets, in which investors struggle to process the overwhelming amount of information available to them[44]. An excessive information overload correlates with decreased trading volumes and predicts higher market returns for up to 18 months. The data suggests that when inundated with information, investors' limited attention spans lead to less trading activity and a demand for higher risk premiums, especially for small-cap, high-volatility and less profitable stocks.

As already discussed, **virality** is probably the main peculiarity of the social media content and it acts as a catalyst for financial decision-making through mechanisms deeply rooted in emotional arousal and psychological contagion. To understand why content goes viral, Berger and Milkman

[42] Shiller, R. J. (2019). *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press.
[43] Simon, Herbert A. (1971). *Designing Organizations for an Information-rich World*. In M. Greenberger (Ed.), Computers, Communications, and the Public Interest, Baltimore, MD: Johns Hopkins University Press, 37-52.
[44] Bernales, A., Valenzuela, M., & Zer, I. (2023). *Effects of Information Overload on Financial Markets: How Much Is Too Much?* International Finance Discussion Papers, No. 1372. Board of Governors of the Federal Reserve System.

conducted an interesting analysis[45] which demonstrates that high-arousal emotions, such as excitement, anger or anxiety, significantly increase the likelihood of content sharing.

Table 2 - Percentage change in the probability of making the most-emailed list due to a one standard deviation increase in various article traits (source: Berger, J., & Milkman, K. L. (2013) [46])

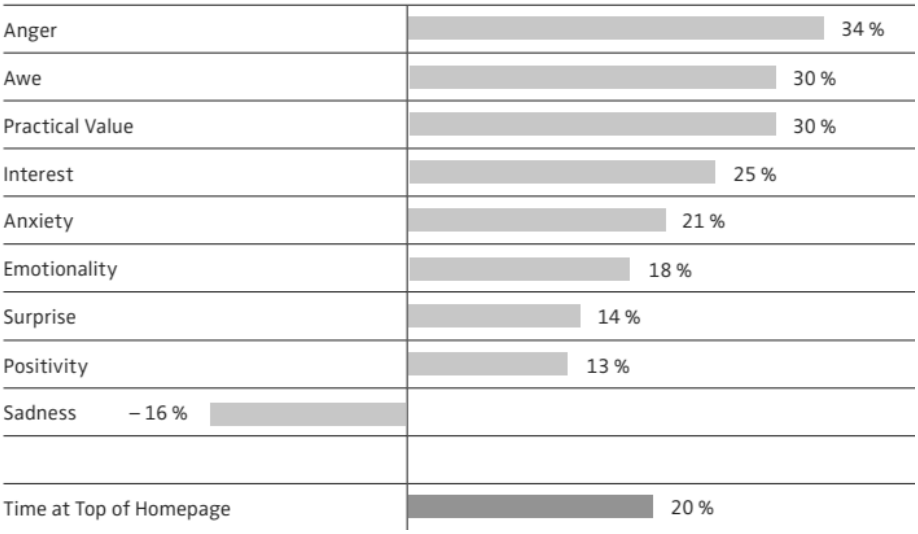


Table 2 clearly shows how deactivating emotions like sadness are negatively correlated to virality, while high-engaging emotions, both negative and positive, such as awe and anger, are positively linked to virality. It's obvious that even in content virality, alongside the actual usefulness of the content itself, the emotive component plays an important part.

Mechanisms such as hashtags, trending topics, and upvoting help amplify certain messages, making it possible for opinions, rumours, or emotional reactions to reach viral status within hours. This amplification is also related to **FOMO**. The fear of missing out emerges as a critical driver in investment contexts, particularly on platforms optimised for rapid engagement. An interesting juncture between social, finance and FOMO is the case of trading platforms like eToro that are organically connected to various social networks enabling their members to share their positions online. This platform's precise strategy is to merge the trading and the social, and they do so by allowing users to follow easily financial instruments and traders they like, interact with other users, and engage in discussions on various topics. For a new member, it's incredibly easy to follow and replicate other user strategies, making trading even more accessible to the unexperienced consumer. This e-democratisation of investing and trading could lead to a boyish gamification and

[45] Berger, J., & Milkman, K. L. (2012). *What Makes Online Content Go Viral?* Journal of Marketing Research, 49(2), 192–205.
[46] Berger, J., & Milkman, K. L. (2013). *Emotion and Virality: What Makes Online Content Go Viral?* GfK Marketing Intelligence Review, 5(1), 18–23.

a betting style of trading which is easy to use and turns investments from a knowledge-based activity into a mere gambling wager[47].

A defining example of how social media reshaped market dynamics is the **GameStop** short squeeze of early 2021. While this event will be examined in greater detail in Chapter 3, a brief analysis will also be provided in this section. The roots of the episode lie in r/WallStreetBets, a subreddit known for its mix of aggressive trading strategies, bold humorous tone and high-risk tolerance. By late 2020, users on the forum began circulating posts pointing out that GameStop (GME) was among the most heavily shorted stocks on the market, with short interest surpassing the number of shares available for trading (approximately 140%). These posts initially took the form of value arguments, mixed with memes and provocations, portraying GME as a classic "underdog" opportunity. However, what turned this from an opposing niche into a market phenomenon was the virality of the content on Reddit. As previously observed, content that evokes high-intensity emotions is significantly more likely to go viral, and this was clearly at play on r/WallStreetBets: GME's posts were often infused with humour, anger at hedge funds and community pride, which created a potent cocktail for mass engagement. Upvotes and comment threads served as social validation, reinforcing the credibility of otherwise speculative claims and creating an echo chamber effect that intensified user conviction. At the heart of this episode was also a narrative (remembering the Shiller's idea): a decentralised band of retail investors fighting back against. In this case, the David vs Goliath story powerfully resonated across digital channels and spilt over into mainstream media, further amplifying the price action. The collective belief system solidified through repetition and mimicry, in a typical herding behaviour, intensified by platform algorithms and emotional validation loops. The situation culminated in January 2021, when GME's share price went from less than $20 to over $400 within a few days. Trading platforms such as Robinhood temporarily restricted purchases of GME, sparking outrage and congressional hearings. Although the price eventually collapsed, the episode raised profound questions about market structure, regulation and the power of coordinated retail investors in the digital age.

The GameStop saga exemplifies how narrative economics and digital virality converge to generate market phenomena that defy traditional valuation logic: to better understand these emotional mechanisms, the next chapter delves into how digital sentiment can be measured, tracked, and even used to anticipate price dynamics.

---

[47] Ivantchev, B., & Ivantcheva, M. (2023). *FOMO effect: social media and online traders*. Journal of Management and Financial Sciences, (48), 59–74.

2.3 Market sentiment and social media

Market sentiment represents the aggregate psychological attitude or collective mood of investors toward financial assets or the market as a whole. Unlike fundamental analysis, which focuses on quantifiable metrics such as earnings, revenue, and financial indicators, sentiment reflects the perception and emotional disposition of market participants rather than economic reality. This distinction is crucial because markets often react to how investors feel about information rather than the information itself.

There are several ways to quantify investor attention on financial markets. **Surveys** represent one of the oldest approaches to measuring sentiment. The American Association of Individual Investors (AAII) Sentiment Survey, for example, asks individual investors whether they are bullish, bearish, or neutral on the stock market over the next six months. Apart from direct surveys and investment newsletters, there also exist some notorious indexes, based on surveys, such as the University of Michigan Consumer Sentiment Index, the Conference Board Consumer Confidence Index, and the UBS/Gallup Index of Investor Optimism. Nevertheless, utilising these metrics might not be the most accurate choice, as these measures are only available monthly or quarterly[48] and, most importantly, there is often little incentive to answer survey questions judiciously or truthfully, especially when questions are highly sensitive[49].

Another approach is to use **financial market-based proxies**. Different studies[50][51]argue that trading volume can effectively serve as a sentiment indicator, as elevated trading activity often corresponds with heightened investor interest; others[52] show that extreme one-day returns capture investor attention, with noise traders typically chasing high-return stocks while avoiding low-performing ones. Baker and Wurgler[53] advocate the use of the CBOE Volatility Index (VIX), commonly referred to as the "fear index", which measures expected market volatility implied by S&P 500 index options: high VIX readings typically indicate elevated fear in the market. However, these market-based indicators can be influenced by factors other than sentiment, such as hedging activities by institutional investors, limiting their interpretive value; moreover, some indicators can

[48] Da, Z., Engelberg, J., & Gao, P. (2015). *The sum of all FEARS: Investor sentiment and asset prices*. Review of Financial Studies, 28(1), 1–32.
[49] Singer, E. (2002). *The use of incentives to reduce nonresponse in household surveys*. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), Survey Nonresponse, Wiley, 163–177.
[50] Gervais, S., Kaniel, R., & Mingelgrin, D. (2001). *The high-volume return premium*. Journal of Finance, 56(3), 877–919.
[51] Hou, K., Peng, L., & Xiong, W. (2009). *A tale of two anomalies: The implications of investor attention for price and earnings momentum*.
[52] Barber, B. M., & Odean, T. (2008). *All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors*. Review of Financial Studies, 21(2), 785–818.
[53] Baker, M., & Wurgler, J. (2007). *Investor sentiment in the stock market*. Journal of Economic Perspectives, 21(2), 129–152.

work pro-cyclical. A classic example of pro-cyclical market sentiment can be observed in the way trading volume attracts investor attention and creates a feedback loop. To better understand this concept, an example will be formulated: imagine a relatively unknown technology stock that suddenly experiences a surge in trading volume, perhaps triggered by a speculative tweet or a viral post on Reddit. The surge appears on stock screens or trend dashboards, attracting the attention of retail investors looking for hot stocks. Intrigued by the unusual activity, other traders jump in, thinking something big is happening. This increased attention leads to more buying, which further inflates the volume, attracting even more participants. In a short time, the share price rises rapidly, not because of new fundamental information, but simply because of the self-perpetuating cycle of visibility, attention and speculation. In these cases, high trading volume becomes both a signal and a catalyst, demonstrating how sentiment-driven behaviour can detach prices from intrinsic values.

A promising and increasingly explored proxy for investor attention is **Internet search behaviour**, based on Simon's theory that decision-making begins with information gathering. Since 2004, tools such as Google Trends (GT) have enabled researchers to measure search volumes for financial terms and activities. Numerous studies[54][55] confirm that peaks in research volume are correlated with future stock returns, realised volatility, and trading activity. Overall, Internet search metrics provide a low-cost, real-time window into retail investor sentiment and attention. Potential drawbacks of this approach are noise (not all research reflects the intent or financial interest of a true investor), lack of contest, and demographic and access bias, as Internet research data potentially under-represents institutional investors or older demographic groups.

The dawn of social media platforms has revolutionised sentiment analysis in financial markets, creating unparalleled opportunities for real-time monitoring of investor opinions. Platforms such as Twitter, Reddit, StockTwits and various financial forums have emerged as vital sources of investor sentiment, where millions of users instantly share thoughts, analysis and reactions to market events. This change has been eased by significant improvements in **natural language processing** (NLP) and machine learning techniques that can now efficiently analyse large amounts of textual data. Modern financial sentiment analysis employs sophisticated methodologies ranging from lexicon-based data mining techniques to traditional machine learning, deep learning and even hybrid approaches[56]. These methods have evolved from fundamental word counters to complex

[54] Preis, T., Moat, H. S., & Stanley, H. E. (2013). *Quantifying Trading Behavior in Financial Markets Using Google Trends*. Scientific Reports, 3, 1684.
[55] Curme, C., Preis, T., Stanley, H. E., & Moat, H. S. (2014). *Quantifying the semantics of search behavior before stock market moves.* Proceedings of the National Academy of Sciences, 111(32), 11600–11605.
[56] Kumar, S., Roy, P. P., Dogra, D. P., & Kim, B.-G. (2023). *A comprehensive review on sentiment analysis: Tasks, approaches and applications* Journal of Economic Surveys, 37(2), 444–478.

systems capable of extracting nuanced features from diverse sources, including news articles, social media posts and financial information.

A pioneering example of how impactful this new era of sentiment-driven finance could be is the study by Bollen, Mao and Zeng (2011)[57], which showed that Twitter moods can significantly predict movements in the Dow Jones Industrial Average. Analysing millions of tweets with the help of sentiment classifiers such as OpinionFinder and Google-Profile of Mood States (GPOMS), the authors found that specific dimensions of mood, in particular 'calmness', had strong predictive power on market trends. Their model achieved an outstanding 87.6% accuracy in predicting daily directional changes in the DJIA, providing one of the first empirical validations of collective mood as a market signal. This study laid the foundation for a growing literature linking social media sentiment to financial volatility, returns and trading activity.

Although these advances in sentiment analysis represent a crucial evolution in financial research and practice, the idyll of perfect real-time mood tracking is still a long way off, as significant challenges remain, including the difficulty of interpreting sarcasm or irony, demographic bias in platform usage, and the volatility of online discourse itself.

2.4 The impact of finfluencers on the market

While the previous section explored how aggregated online sentiment serves to index the mood and attention of investors, it is equally important to examine the sources that generate and amplify such sentiment, as the dissemination of information is not necessarily peer-to-peer. In this context, finfluencers have emerged as key players within the social media ecosystem, not merely reflecting sentiment but actively shaping it. Their ability to elicit widespread engagement, particularly among retail investors, raises critical questions about their influence on market stability. In addition, it must be considered that if sentiment spreads virally, trading patterns may also spread, often resulting in price distortions, liquidity surges and increased volatility. The next section will delve into the mechanisms through which finfluencers influence market dynamics, offering empirical observations and behavioural interpretations of this new phenomenon.

Finfluencers, as mentioned earlier, are content creators specialising in financial topics, sharing investment advice, market commentary and, more generally, financial education pills via social media platforms. They represent the living intersection between the influence of social media and the financial markets, democratising access to financial knowledge while posing new challenges

[57] Bollen, J., Mao, H., & Zeng, X. (2011). *Twitter mood predicts the stock market.* Journal of Computational Science, 2(1), 1–8.

to market stability. Unlike traditional financial advisors, finfluencers typically engage audiences through simplified, accessible and often personalised content, which has particular resonance among younger investors[58]. As observed earlier in the eToro example, the emerging trend of 'social trading', i.e. the tendency to publicly share one's investments and allow other users to replicate them, has made the financial environment even more conducive to the advent of finfluencers. They mainly operate on platforms such as YouTube, TikTok, Instagram, Twitter/X and Reddit, using various content formats from short videos to in-depth analysis posts. Longer video content tends to go to YouTube, shorter content to Tik Tok and Instagram, and text content to Twitter/X and Reddit. Their business models often combine standard platform revenues with sponsored content and affiliate marketing. Research indicates that the appeal of finfluencers stems largely from their perceived authenticity and relationality, qualities that traditional financial institutions often struggle to convey.

Figure 5 - Typical decision-making journey of an average follower (source: Singh, A., & Sarva, R. (2024). The Rise of Finfluencers[59]))

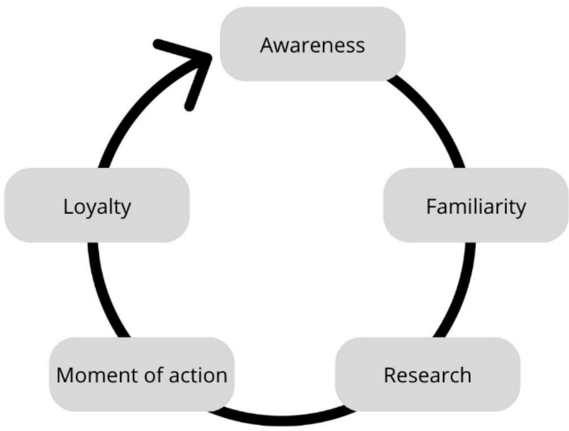


Figure 5 shows the decision path of investors in five stages. In the awareness phase, potential investors are hooked by a specific trigger cleverly positioned by the influencer to capture their attention. Then, in the familiarity phase, the viewer assesses the credibility of the speaker based on trust and perceived affinity mechanisms, often assessed through metrics such as number of followers, engagement rate and community interaction. In the subsequent research phase, the individual seeks further information, either through online resources or by consulting peers.

[58] Hayes, A. S., & Ben-Shmuel, A. T. (2024). *Under the finfluence: Financial influencers, economic meaning-making and the financialization of digital life.* Economy and Society, 53(3), 478–503.
[59] Singh, A., & Sarva, R. (2024). The rise of finfluencers: Financial influencers and the digital investor journey. *Australasian Accounting, Business and Finance Journal*, 18(3), 268–278.

Although influencers' endorsements can increase awareness, studies[60] show that users place more trust in familiar sources, such as friends or family. For many young or inexperienced investors, future speculative returns outweigh past experiences and decisions are often driven by peer pressure or FOMO. Finally, in the moment of action, the follower decides to invest and may even decide to join the influencer community by paying a kind of subscription fee. In the last phase, that of loyalty, the follower becomes an active marketing agent for the influencer by engaging in online word-of-mouth, which leads to a potential expansion of the influencer's popularity.

As far as the actual impact of finfluencers on the market is concerned, the current state of the art is not conclusive. The previously mentioned study by Singh (2024) argues that there is no evidence to state that the activity of finfluencers causes substantial and constant changes to the market. An earlier study by Oosting (2022)[61] conducted specifically on YouTube finfluencers also firmly clarified that their actions do not lead to significant impacts on the market.

It is necessary to point out, however, that the finfluencers are a worldwide phenomenon, which is spreading very rapidly, and that, precisely because of this, there are still no conclusive studies that have managed to sample a sufficiently large number of content creators to be able to make a conclusive judgement. It is also essential to remember that the growth in popularity of social media is exponential and that, in this area, a study conducted even only two years ago could be considered obsolete.

However, it is different when it comes to copy-trading, since research[62] reveals additional liquidity concerns, as automated mirroring of finfluencer trades can create synchronised order flows that amplify price impacts beyond what individual trader actions would generate. This synchronisation creates complex feedback loops between content, sentiment, and market activity.

What is certain, though, is the ability of these content creators to condition investors' choices. As observed in Section 1.3, in fact, younger people in particular are strongly influenced by social networks, both because they are digital natives and because they are usually more open to novelty. Having ascertained this capacity of finfluencers, it is therefore necessary to give serious thought to their **regulation**.

---

[60] Cooley, D., & Parks-Yancy, R. (2019) *The Effect of Social Media on Perceived Information Credibility and Decision Making.* Journal of Internet Commerce, 18(3), 249-269.

[61] Oosting, J. (2022). *Finfluencers and their impact on the stock market* (Bachelor's thesis, Leiden Institute of Advanced Computer Science, LIACS, Leiden University).

[62] Apesteguia, J., Oechssler, J., & Weidenholzer, S. (2019). *Copy trading.* Management Science 66(12):5608-5622.

In this regard, it should be noted that there is a distinction between licensed financial professionals who create content and unregistered influencers who operate outside the regulatory frameworks. The majority of popular finfluencers do not have formal financial credentials or licences, which raises concerns about the quality and reliability of the information disseminated. Complicating the picture, a recent study of finfluencers in the UK pointed out that disclaimers that should be used to warn viewers of the potential risks they might face were found in less than 2% of the videos analysed.

A recent study[63] compared UK, EU and US regulations, which differ, for example, in the definition of investment recommendations. This creates confusion and limits the effectiveness of the regulatory framework for finfluencer content, as such content may meet the regulatory criteria associated with recommendations/advice in some states, while it may not be suitable in others. Furthermore, this analysis provides recommendations for regulators, such as the implementation of a more universal definition of an investment recommendation, increased communication between regulators and content creators, and public recording and reporting of complaints data.

2.5 Brief outlines on socio-physics approaches to behavioural finance

As behavioural finance has increasingly adopted interdisciplinary frameworks to better understand collective investor behaviour, a growing body of literature has looked to **sociophysics**, a branch of statistical physics applied to social systems for theoretical and empirical insights. The term is defined by Săvoiu and Iorga Simăn[64] as follows: "*Sociophysics can be described as the sum of activities of searching for fundamental laws and principles that characterize human behavior and result in collective social phenomena*. "At the heart of sociophysics lies the idea that macroscopic patterns in social systems can emerge from microscopic interactions among heterogeneous agents, much like the dynamics observed in physical systems made of particles. While human decision-making is vastly more complex than atomic interactions, simplified agent-based models and stochastic processes can approximate aspects of herding, opinion dynamics, and contagion that frequently characterise financial markets.

The most promising fields of sociophysical research are Opinion Dynamics, Group Decision Processes and Knowledge Dynamics[65]. These branches, through the application of the principles of Agent-Based Modelling and Social Network Analysis, describe how opinion, consensus and

63 Espeute, S., & Preece, R. (2024). *The Finfluencer Appeal: Investing in the Age of Social Media*. CFA Institute: Industry Future Series.
64 Săvoiu, G., & Iorga Simăn, I. (2013). *Sociophysics: A new science or a new domain for physicists in a modern university*. In G. Săvoiu (Ed.), Econophysics, San Diego, CA: Academic Press, 149–168.
65 Tsintsaris, D.; Tsompanoglou, M.; Ioannidis, E.(2024) *Dynamics of Social Influence and Knowledge in Networks: Sociophysics Models and Applications in Social Trading*, Behavioral Finance and Business. Mathematics , *12*, 1141

knowledge respectively diffuse in various networked social environments. In the information society, one of the social spaces of greatest relevance and impact is that constituted by the interactions between financial agents, and therefore **econophysics**, i.e. the sub-category of sociophysics based on economics that models the behaviour of a not entirely rational agent and the influence that interactions between financial agents have on the prioritisation of assets, is of great importance.

Building on these conceptual foundations, one of the most promising applications of sociophysics, particularly of Opinion Dynamics (OD), concerns behavioural finance and social trading. OD models are used to simulate how financial social agents, such as retail investors, traders, firms and even finfluencers, interact and influence each other within complex networks, thereby shaping market sentiment and price formation[66]. The empirical relevance of these models is reinforced by previous sociological studies that provide strong evidence of the diffusion of financial interest and information among market participants. For example, Shiller and Pound (1989)[67] found that individual investors often rely more on peer recommendations and personal suggestions than on fundamental analysis, while Shive (2010)[68] showed that investor behaviour can propagate through networks in an epidemic-like pattern, leading to large-scale synchronised trading movements.

Importantly, these findings also challenge the assumption of completely rational investors and highlight the Simonian bounded rationality of financial agents. Agents do not operate in isolation, but are deeply embedded in social environments where mimicry, peer validation and emotional contagion significantly influence decision-making. This social embeddedness often leads to short-term behaviour, in which immediate gains and crowd tendencies are privileged over long-term fundamentals[69]. However, despite these cognitive and social limitations, such behaviour can be effectively captured through agent-based models, which simulate individual decision-making and its aggregate effects. The strength of sociophysics and its economic offshoot, econophysics, lies precisely in this modelling power: it allows researchers to represent irrational but structured behaviours in a mathematically consistent framework, enabling a better understanding of how collective phenomena such as herding, volatility clustering and sentiment-driven bubbles emerge in real markets.

---

[66] Khashanah, K., & Alsulaiman, T. (2016). *Network Theory and Behavioral Finance in a Heterogeneous Market Environment*. Stevens Institute of Technology, Financial Engineering.
[67] Shiller, R. J., & Pound, J. (1989). *Survey Evidence on Diffusion of Interest and Information Among Investors*. Journal of Economic Behavior and Organization, 12, 47–66.
[68] Shive, S. (2010). *An Epidemic Model of Investor Behavior*. Journal of Financial and Quantitative Analysis, 45(1), 169–198.
[69] Oldham, M. (2019). *Understanding how short-termism and a dynamic investor network affects investor returns: An agent-based perspective*. Complexity, 2019

3.ANALYSIS

3.1 Comparative analysis of sources

As stated in the aims of the thesis, the main ambition of this paper is to understand how social media platforms and financial influencers reshape investor behaviour and contribute to changes in the volatility and liquidity of financial markets. This chapter synthesises and critically compares the main bodies of literature presented earlier, bringing together behavioural finance theory, sentiment analysis and sociophysics-inspired modelling, to assess whether these frameworks can sufficiently capture the complexity of digital market dynamics. Rather than recapitulating previous content, this section integrates ideas and offers a critical response to our guiding research question: how do social networks and finfluencers influence financial decision-making, and can existing academic models adequately interpret their influence?

One of the clearest themes that emerged from the literature review is the amplification of classic behavioural biases in digital contexts. As discussed in Section 2.1, fundamental studies on behavioural finance have long identified the tendency of investors to act irrationally under the influence of heuristics such as overconfidence, loss aversion and herding. However, in the age of smartphones and algorithm-curated content, these cognitive shortcuts are no longer just internal flaws but become socially contagious: the habitual use of smartphones for trading activities not only reinforces overconfidence and herd reactions, but also lowers the threshold of decision-making, creating a silent behavioural epidemic. Although these trends are not new, their digital acceleration introduces new levels of complexity that traditional models, rooted in assumptions of individualistic and deliberate decision-making, fail to anticipate.

As behavioural biases are amplified by digital interactions, the architecture of the platforms themselves adds a crucial level of complexity. Social networks are not simply neutral containers of communication: their design actively encourages emotional engagement and the rapid dissemination of information. As discussed in Section 2.2, these structural dynamic fosters what Shiller called ‘narrative contagion’: stories that are compelling not because they are true, but because they are easy to repeat and have great emotional resonance. When applied to financial contexts, this contagion leads to the proliferation of simplified investment theses, which are often detached from the underlying fundamentals but retain a high degree of virality. Although behavioural finance theory provides a solid basis for understanding these heuristics in individual agents, it requires reinforcement through models that take into account **collective behaviour** and amplification of the digital signal. Herding, for instance, takes on a new character in the context of social media. As discussed above in relation to Scharfstein and Stein's (1990) reputational herding, the classical assumption is that individuals imitate others to avoid the reputational cost of standing out. But in online forums, especially anonymous or pseudonymous ones like Reddit's r/WallStreetBets, the reputational calculation is reversed. Bold and unconventional bets are not only accepted, but often celebrated, and upvotes, likes and reposts create a system of social validation that replaces traditional financial incentives. Thus, digital herding is less about conformity and more about performativity. This distinction is crucial because it shifts the psychological factors that determine market coordination and requires a theoretical adjustment in our understanding of how online groups coalesce around financial decisions.

Another aspect that needs to be revisited is the timing of information transfer. In traditional settings, financial information cascaded from analysts to institutional investors and finally to the wider public. Today, this hierarchy has been flattened. A Twitter account or a TikTok profile is all that is needed to become a broadcaster of financial opinion. This “flattening” speeds up market responses so that they may be compressed to just minutes or even moments. Things that would have been diffused slowly, like earnings misses or metal M&A speculation, are immediately absorbed, refracted through the fun-house mirror of digital communities.

Since market efficiency theory (and specifically in its semi-strong form) assumes that prices incorporate all available public information, it is not as evident how this applies to information that is not only accessible, but emotionally intense and algorithmically circulated. In modern times, especially with finfluencers, there has been a rise in a new intermediary: someone who does not filter or analyse information in the traditional sense but rather packages it in such a way as to have the maximum impact on their audience. These s blur the distinction between teacher and entertainer, oversimplifying risk into short, easily shareable stories.

The rise of finfluencers also raises deeper concerns. As analysed by Hayes and Ben-Shmuel (2024), financial influencers do not merely reflect investor interest but actively shape it through the narratives and symbols used in their campaigns. Thus, their audience does not just passively receive advice but engages emotionally and socially with a persona that often, paradoxically, appears more authentic than institutional voices. This introduces what psychology calls “parasocial relationships,” the phenomenon whereby followers are able to develop a sense of trust and familiarity with someone they do not know personally. In financial contexts, this dynamic can be extremely powerful. Unlike traditional advisors, who are required by law to prioritize their clients and follow regulatory guidelines, finfluencers operate in a relatively unregulated space. While their democracy-based potential is undeniable, they also expose followers to narratives that can promote speculative behaviour, precisely because they are outside of regulations.

Current models of rational financial decision making do not account for this parasocial influence, as emotional as it is mimetic. The emotional tenor of finfluencers' communication which often comes with calls for urgency or **FOMO** (or moral confrontation, as in "us versus the hedge funds") has real effects on market dynamics, especially in periods of high retail participation.

Returning to the GameStop/Reddit case, it can be sustained that the GME price surge was not driven by new fundamental information, but by the collective belief of a digital crowd. This belief was reinforced by repeated messages over and over again, emotional connection, and a kind of antagonism toward institutional investors. Efficient markets theory, even in its behavioural extensions, struggles to explain the intensity and duration of this anomaly. While market anomalies have long been recognised by behavioural finance, the size and structure of the GameStop phenomenon suggested something more systemic: the formation of "financial flash mobs," organised not by the fundamentals of economic science but by narrative, personalities, and memes. In this sense, the combination of behavioural finance with sociophysics, particularly the dynamics of opinion, offers a promising avenue for future research. These interdisciplinary frameworks allow the experts to model how exchanges of opinion at the micro level also lead to coordination at the macro level, without assuming full rationality or individual information processing.

Moreover, it's crucial to consider how these online-driven behaviours actually affect traditional market mechanics, such as liquidity and volatility. Although spikes in volatility have always occurred, the magnitude and synchrony observed during retail-driven events appear to be correlated with spikes in online activity. This is particularly relevant in the context of copy trading and speculative investment apps, where the trades of finfluencers are rapidly imitated by followers. The synchronisation of these flows can lead to liquidity collapses or, conversely, sudden surges, depending on the direction of sentiment. The problem is no longer just one of individual irrationality, because now it is truly about systemic coordination: when thousands of users act on the same content at the same time, even a relatively small financial base can generate significant price pressure. Traditional microstructure models do not fully anticipate these reflexive cycles between content and execution, suggesting the need for theoretical innovation.

Assessing literature, it is also worth noting that **sentiment analysis** tools have been analysed as one of the few bridges between digital behavioural dynamics and financial modelling. As discussed in the study by Bollen et al. (2011), aggregate mood signals from Twitter have been shown to have predictive power over movements in the Dow Jones Industrial Average. These findings, though discussed, have sparked a body of research that seeks to analyse and exploit online user-generated content to make market predictions. It has now become common practice to process natural language on Reddit, StockTwits and other platforms and develop sentiment indexes that reflect the emotions of retail investors in real time. However, challenges persist. One problem that has remained is distinguishing real signal from noise. Not all viral posts reflect real trading intentions, and not all social activity translates into actual market movements. In addition, the platforms are highly heterogeneous: the tone and user base of Twitter differ significantly from those of Reddit or even TikTok, complicating any attempted effort to generalise sentiment metrics. However, the growing convergence between AI tools and behavioural finance offers hope for more accurate modelling of the impact of digital sentiment on asset prices.

In light of what has been analysed, the logic behind our empirical analysis now becomes clearer. In section 3.2, these concepts will be directly tested by examining the correlation between Reddit sentiment during the GameStop short squeeze and observable market metrics of liquidity and volatility. This approach allows for the isolation of specific moments of spikes in sentiment and follows their translation onto market action. The case was chosen not only because of its notoriety, but also because it offers an experiment that actually happened: a clearly delimited event with high digital coordination, huge trading volume, and significant price movement. Although the event is unique in terms of scale, it's still useful as a paradigmatic example of broader structural changes that are reshaping market behaviour. Regarding the impact of finfluencers, the dissertation does not pursue a broad statistical study, both because of data access limitations and methodological concerns, but rather turn to illustrative cases of meme stocks or speculative tokens whose price paths appear to be closely linked to influencer promotion. These examples underline the need for regulatory oversight and the limitations of current theoretical tools.

In conclusion, our comparative analysis reveals that traditional behavioural finance remains an essential lens, but it must be expanded by always taking into account the interdisciplinary contributions of digital sociology, sentiment analytics, and sociophysics. By now, the fundamental assumptions of individual rationality, efficient information processing and static preference formation are increasingly challenged in an age when content now spreads faster than financial fundamentals and when (media) attention, rather than capital, is often the main driver of price action.

3.2 Quantitative Analysis: Linking Reddit Sentiment to GME Market Dynamics

This section undertakes the quantitative analysis at the heart of this thesis, examining the relationship between collective sentiment on **Reddit** and the behaviour of **GameStop** (GME) stock during the dramatic short squeeze episode in late January 2021. The underlying objective is to go beyond the theoretical discussion and provide empirical evidence for the thesis that digital communities, through their rapid propagation capacity, emotional contagion and coordinated action, can measurably influence market volatility and liquidity, even in ways not predicted by classical financial theory. Focusing on the representative case of GME, this section aims to clarify whether and how social media-driven sentiment manifests itself in market data, thus creating a powerful bridge between the qualitative and quantitative dimensions of behavioural finance.

The dataset[70] used for this analysis is a comprehensive collection of all publicly available comments posted on the **r/WallStreetBets** (WSB) subreddit until 16 February 2021. For further clarification on the dataset, please refer to the paragraph *A.1 Dataset*. The unfiltered and unprocessed data, preserves the authentic, informal and often chaotic nature of the discourse on Reddit. To use this enormous amount of data, coding is necessary. In this case, Python will be used together with specific libraries, as explained in *A.2 Python and libraries*.

To isolate content relevant to GameStop, a targeted keyword filtering process was applied. The use of keywords enables successfully tracking comments targeting a particular issue, company, or stock even in large and unstructured data repositories like social media posts. On the WallStreetBets community of Reddit, the posters will typically reference GameStop by a variety of words, some of which are the ticker symbol ("gme"), brand name ("gamestop"), as well as trendy slang forms like "gamestonk." In using these terms, the filtering algorithm captures not only explicit references to the company, but also the colloquial talking, memes, and in-group proprietary terms.

The filtering was then continued with a case-unsensitive approach, reducing to approximately 171,000 GME-related comments to be analysed in detail. Recognising the informal and eccentric nature of the discourse on Reddit, characterised by heavy use of slang, memes and non-standard spelling, a clean-up protocol was implemented. This protocol systematically removed extraneous content that could interfere with sentiment analysis, including hyperlinks (URLs), ticker symbols and monetary markers (e.g. '$GME'), references to specific users (denoted by 'u/username') and other subreddits ('r/subreddit'), as well as residual HTML entities, i.e. special web codes, that sometimes appear in web-scraped data.

A particularly interesting element of the pre-processing code concerned the handling of emoji. Emoji play a key role in the WSB community, as they often convey intense sentiment, irony or

[70] Podolak, M. (2021). *Reddit WallStreetBets Comments* [Data set]. Kaggle.

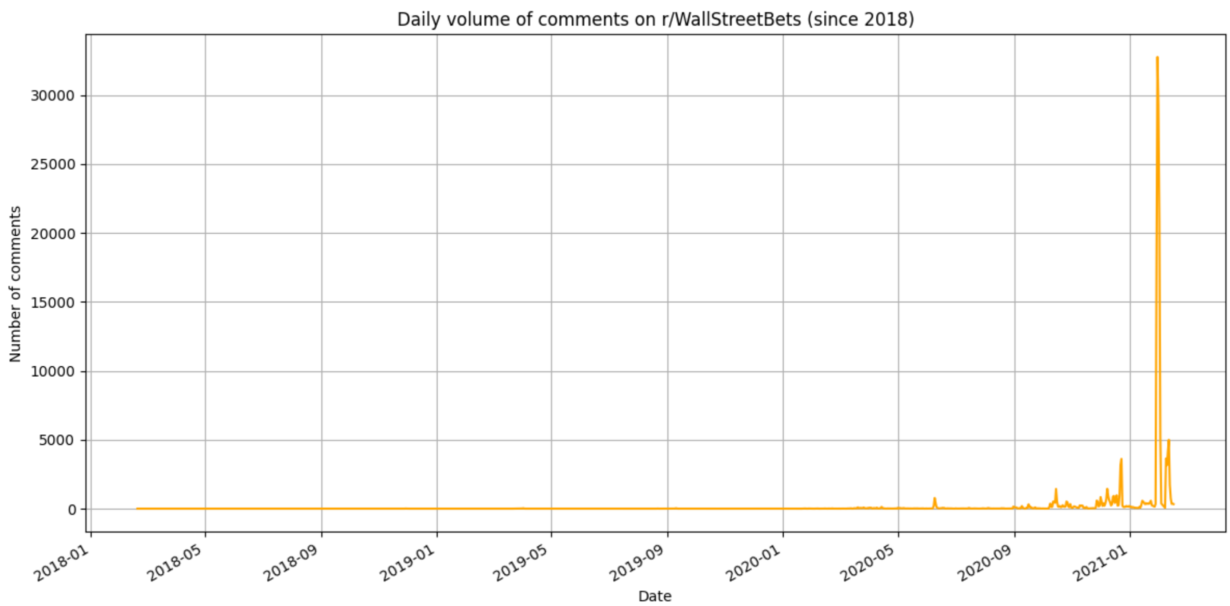
community jokes that are not easily captured through text alone. For these reasons, emojis have been treated with specific codes in the manner indicated in *A.3 Emoji handling*.

Together, these steps produced a dataset of GME-focused Reddit comments that was not only clean and consistent, but also expressive of the underlying sentiment signals, regardless of whether they appeared in conventional text, meme or emoji-rich language. This pre-processing was essential to the validity of the subsequent NLP-based sentiment scoring, allowing the analysis to more accurately quantify the true emotional climate of the WallStreetBets community during the GameStop short squeeze. Natural Language Processing (NLP) refers to computer methods that make it possible to analyse and interpret human language, helping to identify patterns and emotions within large collections of text data like Reddit comments. NLP is applied widely in virtually all industries, from translation and virtual assistants to spam blockers and social media monitoring and relies on a variety of analysis tools and algorithms.

Given the informal language, a sentiment analysis tool suitable for social media was essential. For this reason, the **VADER** algorithm was employed. For a more detailed description of VADER and how it is used, see *A.4 VADER for sentiment analysis*.

To ensure that the subsequent sentiment analysis would be both statistically meaningful and contextually relevant, particular attention was paid to the temporal dynamics of online discussion volume.

Figure 6 - Daily comments volume on r/WallStreetBets



As illustrated in Figure 6, the daily number of GameStop-related comments on r/WallStreetBets remained relatively low and stable for much of the period between 2018 and the end of 2020.

However, in the weeks leading up to and during the January 2021 short squeeze, the volume of discourse increased exponentially, culminating in a dramatic spike that surpassed all previous activity on the subreddit regarding GME.

Given this sharp discontinuity in the data, the analysis was intentionally focused on the period of the short squeeze (December 2020 to February 2021), when both attention and market volatility peaked. Focusing on this interval not only maximises the relevance of the sentiment signal (capturing the full intensity of the crowd-driven narrative dynamics) but also ensures that the analysis is not diluted by periods of relative inattention.

In the quantitative analysis, two complementary approaches were considered to aggregate the daily Reddit sentiment related to GME: an **unweighted** and a **weighted** method. The unweighted approach simply calculates the average sentiment score for all relevant comments each day, treating each comment as equally important.

In order to weight each comment correctly, a hybrid approach was chosen, differentiating those with more upvotes than downvotes (hence positive scores) and those with negative scores. For the former, it is logical to weight the comments with the most upvotes more heavily, but to do so linearly would have resulted in an excessive monopolisation of the final sentiment by these (e.g. a comment with 100 upvotes would have weighed 100 times more than a normal comment). For these reasons, a logarithmic approach was chosen to mitigate the incremental effect of upvotes. Instead, for low score comments, the idea was to penalise comments more severely as the score decreased. This was chosen because very downvoted comments do not reflect the opinions of the community and it is therefore fair that they are worth almost zero in the model, while those with a slightly negative score should still be considered as they are more in line.

When aggregating daily sentiment, each comment i with score score_i receives a weight $w_i$defined as:

$$w_i = \begin{cases} 1 + \log(score_i + 1) & if\ score_i \geq 0 \\ e^{\frac{score_i}{k}} & if\ score_i < 0 \end{cases}$$

for the analysis the decay parameter it's set to $k = 10$, because it penalises strongly downvoted comments. The choice of this parameter is explained in more detail in *A.5 The choice of parameter k for the weighting function.*

The daily **weighted sentiment** for day t is then computed as:

$$WeightedSentiment_t = \frac{\sum_{i=1}^{N_t} w_i \cdot s_i}{\sum_{i=1}^{N_t} w_i}$$

where s_i is the sentiment score (VADER compound) of comment i, and N_t is the number of comments on day t.

To determine which aggregation provided more meaningful insight into market behaviour, both unweighted and weighted daily sentiment series were compared.

Figure 7 - Unweighted vs weighted daily sentiment scores comparison

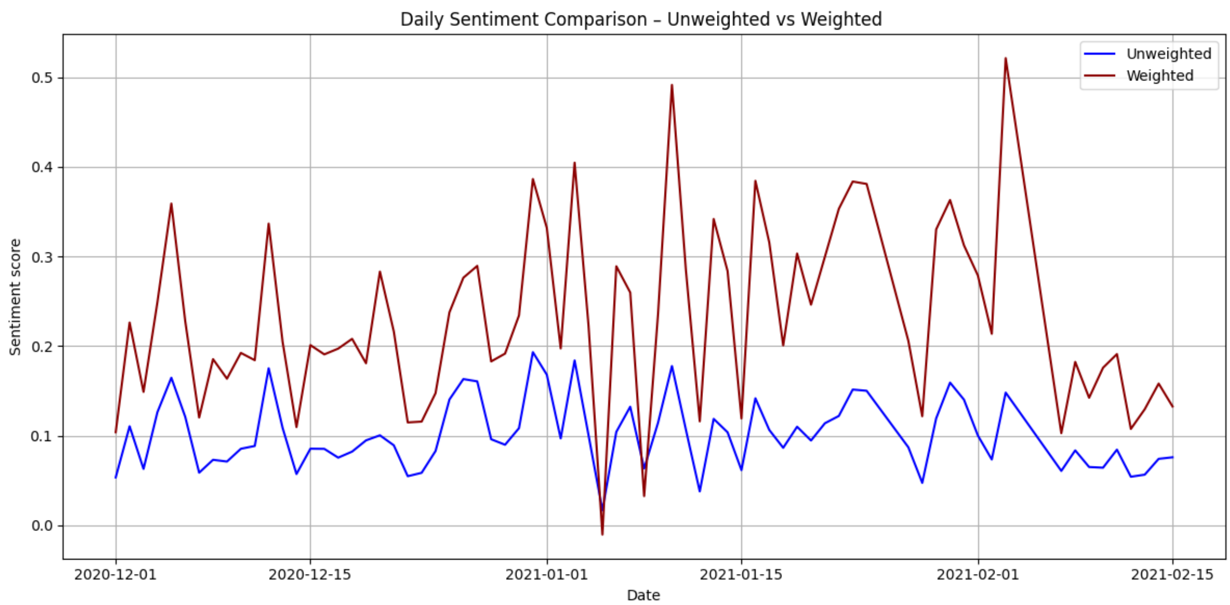


Figure 7 illustrates GME's daily sentiment in the short squeeze period using both the unweighted (blue line) and weighted (red line) methods. The comparison reveals that the weighted approach produces a more volatile sentiment curve, which more closely follows the peaks of attention and engagement.

After this comparative evaluation, the weighted approach was chosen for the following analysis, as it more accurately represents the dynamics of information flow and influence on Reddit, particularly during the stormy short squeeze episode.

To investigate whether shifts in Reddit sentiment translated into observable effects on financial market variables, daily GME market data was methodically collected and aligned with the sentiment time series. Market data was sourced directly from Yahoo Finance (via the yfinance Python library), providing records of GameStop's daily open, high, low, close (OHLC) prices, as well as trading volume. This data enabled a granular analysis at the daily level. In addition to the raw price and volume series, additional variables were computed to provide a more comprehensive picture of market dynamics. Daily returns were calculated as the logarithmic difference of the closing price, and a rolling 7-day realised volatility measure was constructed to quantify short-term price instability.

The daily weighted Reddit sentiment series, constructed as described in the previous section, was directly compared to three key market metrics: closing **price**, trading **volume** (as a proxy for market liquidity), and 7-day realised **volatility**.

Figure 8 - GME closing price and weighted sentiment

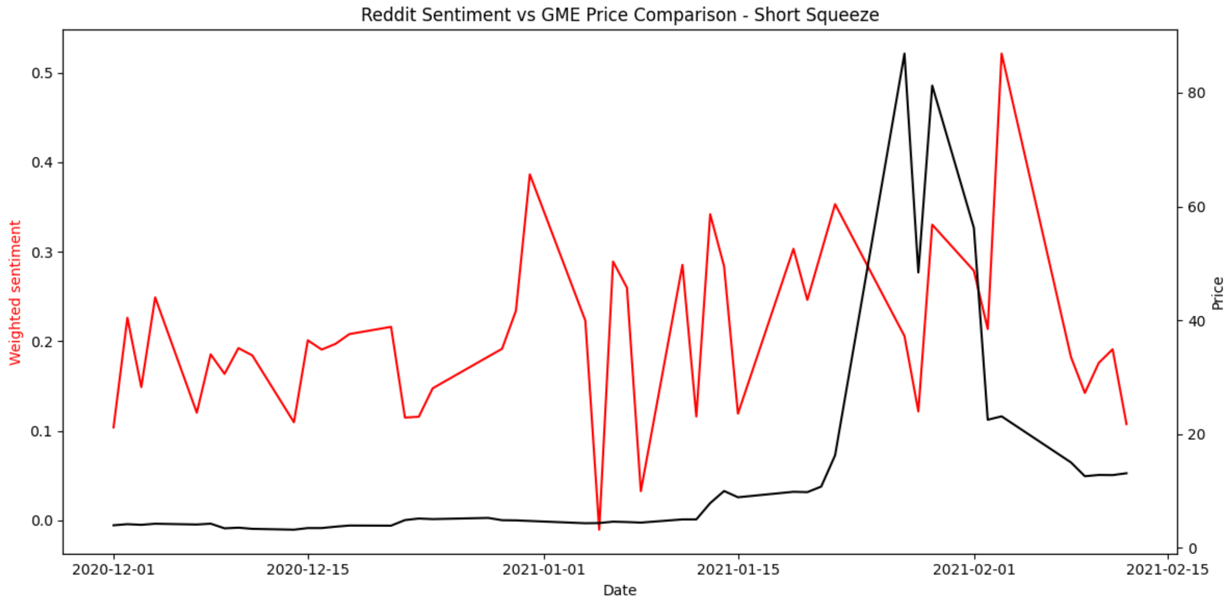


Figure 8 compares the daily weighted Reddit sentiment with the closing price of GME stock during the short squeeze period. The most notable visual feature is the dramatic increase in GME's stock price at the end of January 2021, a movement that is slightly mirrored in the collective sentiment on Reddit, although with different timing and with different intensity. Looking at the two series, there does not appear to be any particular correlation.

Figure 9 - GME Volume and weighted sentiment

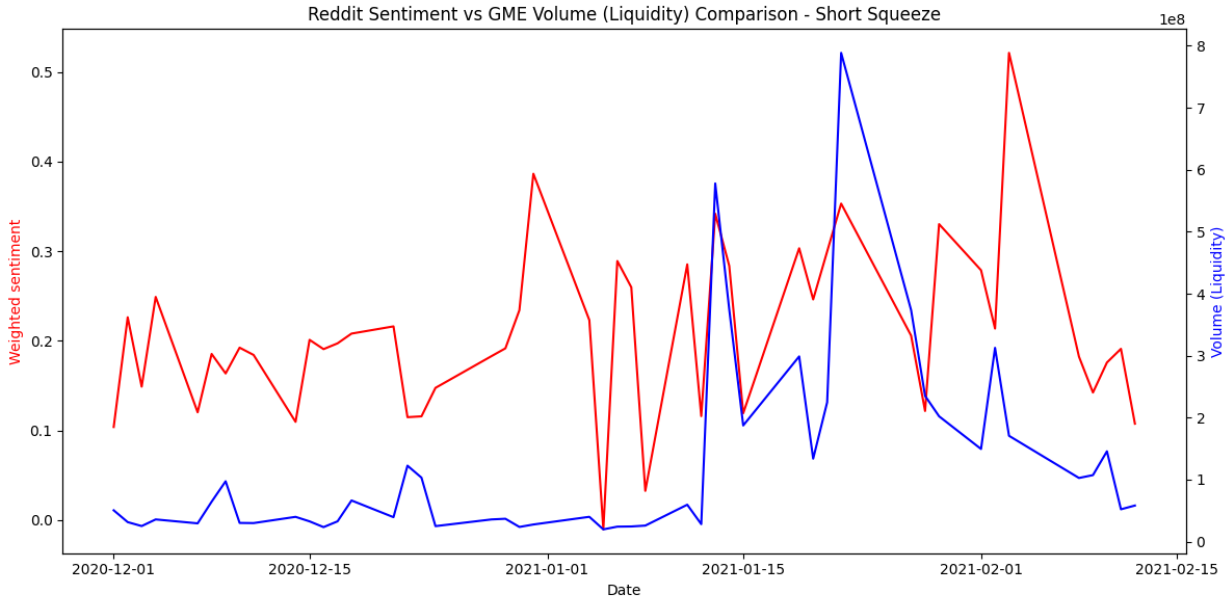


Figure 9 depicts the relationship between the weighted Reddit sentiment and the daily trading volume (a proxy for market liquidity) of GME. It reveals an alignment between spikes in online sentiment and surges in trading activity, particularly during the most intense phases of the short squeeze. While there are certainly some conflicting movements, these similarities in trends may suggest that collective sentiment expressed on social media might have directly contributed to extraordinary levels of trading and liquidity.

Figure 10 - GME Volatility and weighted sentiment

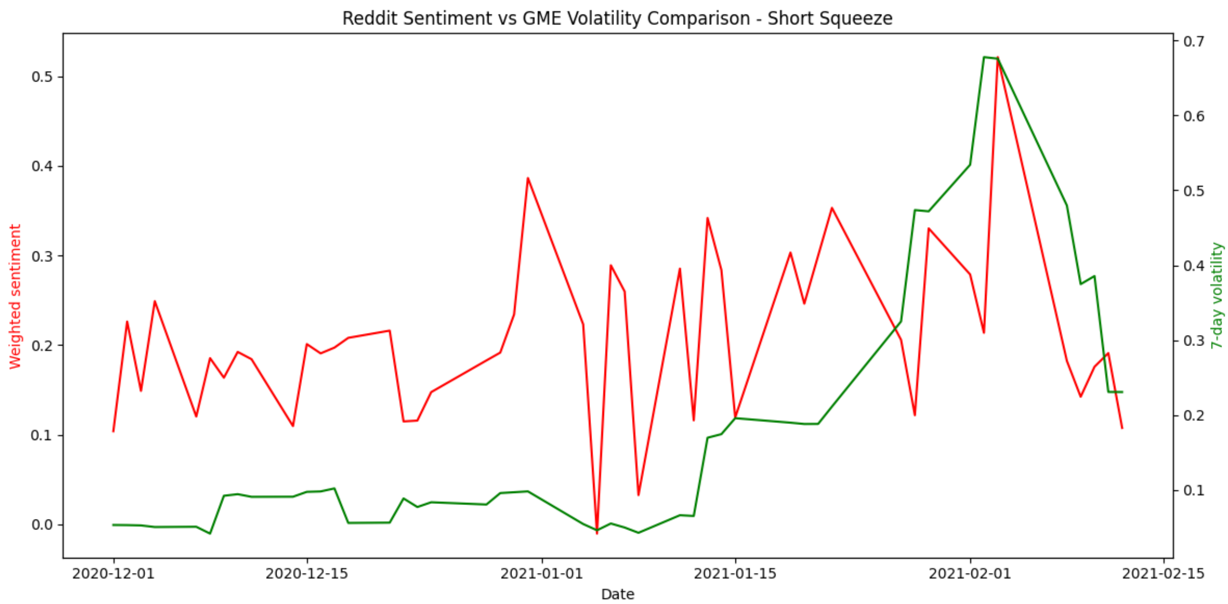


Figure 10 illustrates the co-movement of weighted Reddit sentiment and a 7-day rolling volatility measure for GME. Notably, the increase in online enthusiasm corresponds to periods of greater price instability, highlighting the connection between digital community dynamics and traditional risk metrics. The chart, which appears to be the one with the most evident correlation of the three, highlights how waves of polarised sentiment can amplify volatility during periods of market stress.

A suitable method to assess whether the observed co-movements between sentiment and the main market variables are statistically significant is correlation analysis. The **Pearson correlation** coefficient was applied in this investigation. This coefficient, often denoted as *r*, is a widely adopted statistical measure that quantifies the strength and direction of the linear relationship between two continuous variables. The value of *r* ranges from -1 to +1: a value of +1 indicates a perfect positive linear relationship, -1 indicates a perfect negative linear relationship, and 0 indicates no linear relationship. The Pearson correlation is calculated by dividing the covariance of the two variables by the product of their standard deviations. In practical terms, it measures how much the variables move together relative to how much they vary individually. A coefficient close to +1 means that as one variable increases, the other tends to increase proportionally; a coefficient close to -1 means that as one increases, the other tends to decrease. Values near 0 suggest no consistent linear relationship. It is important to note that the Pearson correlation captures mainly linear relationships and may not reflect non-linear associations between variables.

This was followed by a significance test to determine whether these correlations could plausibly be attributed to chance, using standard p-value thresholds ($p < 0.05$ indicates statistical significance).

Table 3 - Correlations and p-value results

Pair	r	p-value	Significance
Sentiment Weighted vs Volatility (7d)	0.346	0.0171	Statistically significant
Sentiment Weighted vs Volume (Liquidity)	0.414	0.0038	Statistically significant
Sentiment Weighted vs Price (Close)	0.218	0.142	Not significant

The results, summarised in Table 3, show that the correlation between weighted sentiment and trading volume and volatility was positive and statistically significant (correlation with volume: **r = 0.414**, p = 0.0038; correlation with volatility: **r = 0.346**, p = 0.0171). These results provide solid empirical support for the hypothesis that waves of online enthusiasm and crowd sentiment on social media are related to tangible increases in market activity and price instability, and that therefore the mechanisms driving liquidity and volatility in modern markets cannot be understood through classical models of investor rationality and information efficiency alone.

The study obviously has limitations: the main one stems from the use of VADER, a sentiment analysis tool designed for social media but still imperfect because when it comes mainly to nuances such as sarcasm, inside jokes or context-dependent expressions they may not be fully captured. As a result, sentiment scores are best interpreted as rough approximations rather than definitive measures of community mood. Other limitations relate to sampling, as despite the careful choice of keywords some relevant comments may not have been considered. Finally, despite the significant correlation between sentiment and the market proxies analysed, there were insufficient statistical tests to assert cause-and-effect relationships.

Nevertheless, in spite of these limitations, the present study provides solid evidence that the collective sentiment on Reddit can be quantitatively related to the main market dynamics during the GME short squeeze. The results demonstrate that, even in the presence of imperfect data, social media sentiment analysis offers significant insights into the mechanisms through which online communities influence financial markets.

3.3 Brief case study on the finfluencer impact

While the previous section explored how collective sentiment on social media platforms such as Reddit can impact market behaviour, it is equally crucial to consider the direct influence of individual finfluencers on trading dynamics. This complementary analysis aims to illustrate that, alongside the power of online communities, individual voices with immense reach can trigger equally sudden and significant market events. Finfluencers, especially those with celebrity status, can quickly mobilise public attention, create waves of speculative enthusiasm and, sometimes unwittingly, encourage herd behaviour or even market manipulation.

This section briefly examines the case of **Elon Musk** and his well-documented influence on the price of **Dogecoin** (DOGE). As CEO of Tesla and SpaceX, Musk has a large following on Twitter (now X), where his posts routinely reach tens of millions of users in real time. Unlike traditional financial commentators, Musk's social media presence blurs the line between entertainment and investment advice, combining humour, memes and cryptic statements with bold opinions on market trends.

Dogecoin, originally launched in 2013 as a light-hearted parody of the cryptocurrency boom, has quickly transformed from an internet joke into a digital asset with real economic consequences, largely thanks to the wave of attention Musk's own tweets have generated. Although Dogecoin's fundamental value is minimal compared to major cryptocurrencies, its price trajectory has been repeatedly and visibly influenced by Musk's social media interventions. Over the course of 2021, a handful of Musk's tweets caused an unprecedented surge in both price and trading volume, grabbing headlines around the world and spurring speculation about the boundaries between opinion, manipulation and legitimate market influence.

This case is particularly illustrative because it not only highlights the speed and scope of influence enabled by social media but also exposes the regulatory blind spots and major risks that retail investors face in the digital age. By analysing immediate market reactions to a series of Musk's most influential Dogecoin tweets, this case study aims to demonstrate the outsized role that charismatic finfluencers can play in shaping market liquidity, volatility and investor decisions, often in ways that defy conventional economic logic. Three of the Musk's most impactful Dogecoin-related tweets are presented and commented.

Figure 11 - "Dogecoin is the people's crypto" (February 4, 2021) (source: X)

This declaration represents one of the earliest and most direct endorsements of the coin to his millions of followers. The tweet was posted at 9:15 AM UTC and rapidly went viral, amassing hundreds of thousands of likes and retweets within hours. In particular, this simple statement, devoid of any technical analysis or financial advice, triggered a wave of public interest and was immediately picked up by news outlets and social media platforms.

Figure 12 - "Doge Barking at the Moon" (April 15, 2021) (source: X)

Elon tweeted an image accompanied by the phrase "Doge barking at the moon". The image is from Joan Miró's surrealist painting "Dog barking at the moon" (1926) and in this case the pun is about the fact that Dogecoin's iconic mascot Shiba Inu is also a dog, but not only that as the painting also evokes the cryptocurrency meme "to the moon" (signalling the expectation of skyrocketing prices). This tweet is an excellent example of how Musk combines meme culture with artistic references to maximise his online influence.

Figure 13 - "The Dogefather, SNL May 8" (April 28, 2021) (source: X)

The post announces Elon's imminent appearance as host of Saturday Night Live (SNL), a long-running American television programme known for its comedy sketches and huge national and international audience. SNL has long been considered a pop culture institution in the United States, and a host's appearance usually arouses considerable media attention and public anticipation. Referring to himself as “The Dogefather”, a pun combining Dogecoin and “The Godfather”, as well as a self-attributed leadership title, Musk fuelled a wave of speculation that his appearance at SNL would further promote Dogecoin to a mainstream audience.

To conduct this analysis, daily historical data on the price, trading volume and realised volatility of DOGE was collected from Yahoo Finance, for the period between January and June 2021. Volatility is calculated as the 7-day annualised standard deviation of daily returns, a standard measure to capture short-term price instability. The analysis focuses on the three key posts by Elon above commented, each of which received significant attention and can be dated accurately. Each tweet has been placed on the timeline allowing us to visually and quantitatively assess the immediate and short-term market reaction. For each event, three **time-series graphs** are presented: the DOGE price, trading volume and 7-day realised volatility, all annotated with the dates of the tweets for clarity.

Figure 14 - DOGE price and Elon tweets

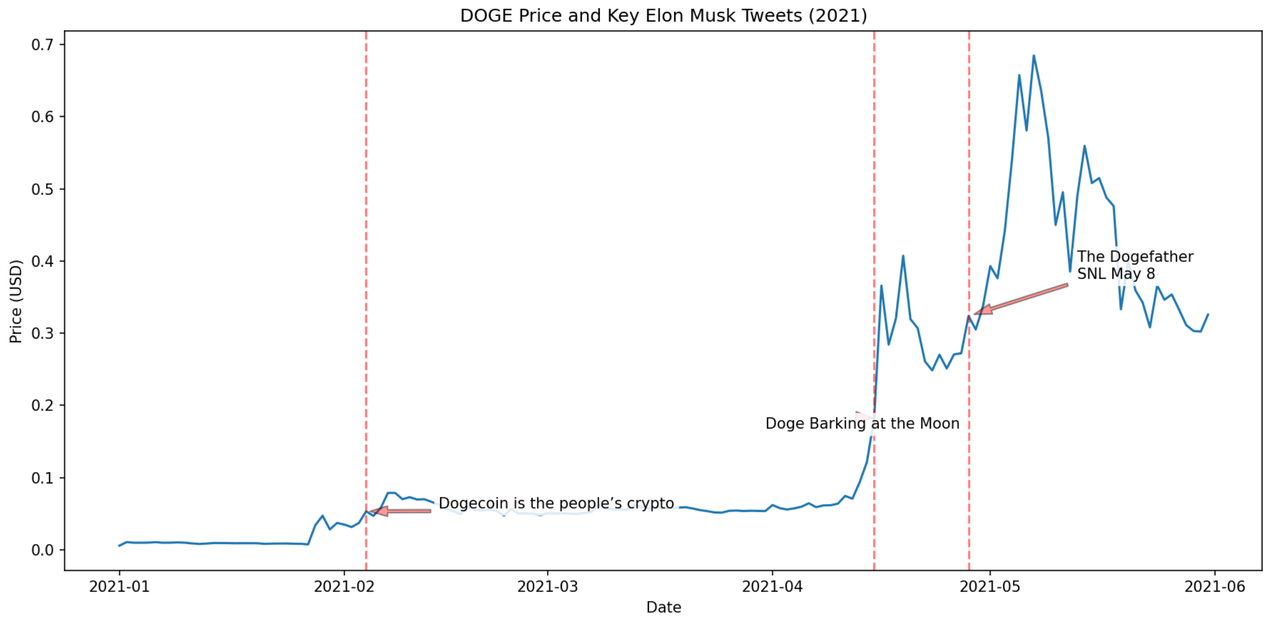


Figure 14 visualises the daily closing price of Dogecoin from January to June 2021, overlaid with the dates of three of Elon Musk’s most influential tweets. It is immediately apparent that each major tweet is followed by a sharp and rapid surge in DOGE’s price. Specifically, the barking at the moon one was already in a bullish period and the tweet clearly amplified it.

Figure 15 - DOGE trading volume and Elon tweets

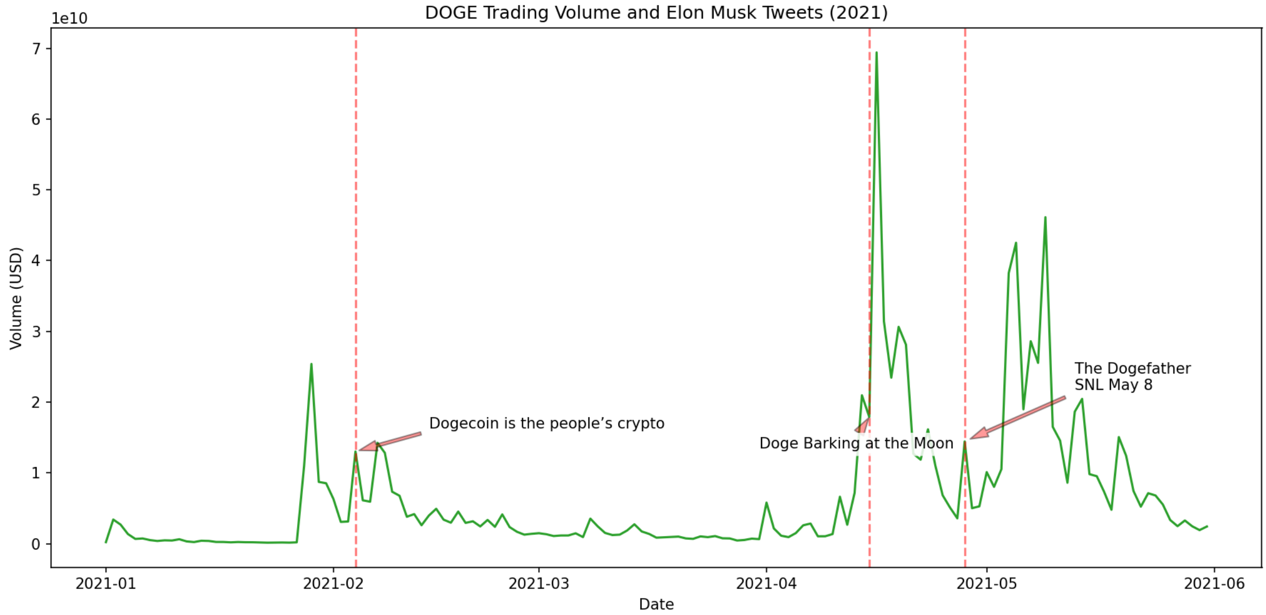


Figure 15 shows the daily trading volume of DOGE in USD over the same period, re-marking the three selected tweets. Apart from the first tweet, which did not have a big spike in volume (probably because the cryptocurrency had already spiked significantly before for other reasons), the others had a massive impact, as the graph clearly shows.

Figure 16 - DOGE volatility and Elon tweets

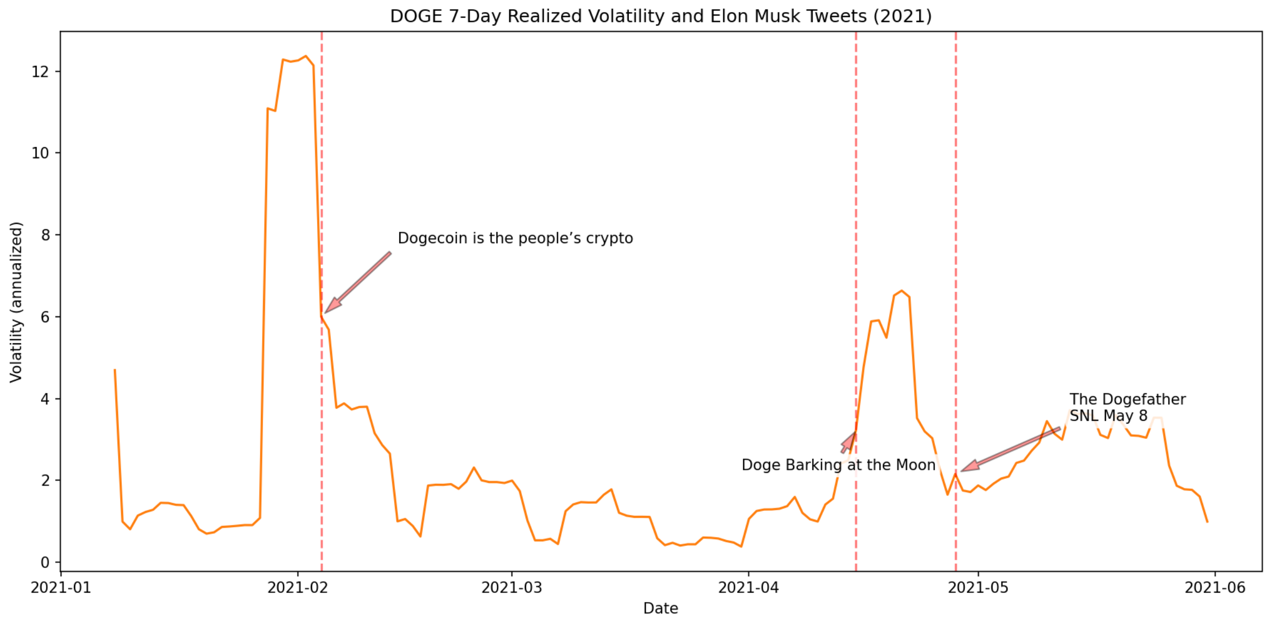


Figure 16 is similar to the volume chart, as expected. Before the first tweet there is a big spike in volatility because the coin had already been the subject of much discussion, precisely on r/Wallstreetbets, which as has been analysed above, was experiencing the peak of popularity at that time, which led to DOGE getting a lot of attention.

Table 4 - DOGE market reaction after Elon tweets

Tweet Date	Tweet Content	Price Change	24h Volume Change	7-Day Volatility Change*
2021-02-04	“Dogecoin is the people’s crypto”	↑ from \$0.053 to \$0.058 (+8%) over 2 days	↓ from \$13.0B to \$5.9B (–54%)	↓ from 6.0 to 3.8 (–37%)
2021-04-15	“Doge Barking at the Moon”	↑ from \$0.182 to \$0.320 (+76%) over 3 days	↑ from \$17.9B to \$23.5B (+31%)	↑ from 3.2 to 5.9 (+84%)
2021-04-28	“The Dogefather, SNL May 8”	↑ from \$0.324 to \$0.685 (+112%) over 9 days	↑ from \$14.5B to \$28.6B (+98%)	↑ from 2.2 to 2.5 (+15%)
* Volatility is measured as the annualised standard deviation of daily log returns (dimensionless ratio, e.g., "6.0" = 600% annualised volatility). This quantifies the expected price fluctuation, scaled to an annual basis, which is standard in financial risk analysis				

Table 4 provides a clear quantitative snapshot of how Elon Musk's tweets caused immediate and dramatic changes in the Dogecoin market, across three key variables: price, trading volume and realised volatility. Each tweet is followed by a marked reaction, underlining the importance of the tweets' impact.

After the first tweet, there was a modest increase in price (+8%), but a significant decrease in volatility and volume. This can be explained by the fact that the increased volatility was already present before the tweet, powered by the Reddit hype; Musk's tweet thus intervened in an already hyper-speculative context, marking almost a post-peak normalisation.

The subsequent tweet is instead associated with a sharp increase in both price (+76%) and volume (+31%) and volatility (+84%). The meme and artistic reference quickly went viral, accentuating the participation and speculation of the crypto audience, and thus showing how the finfluencer social narrative can push all metrics upwards

Even more evident were the reactions to the latest tweet that provoked a maximum upward rush (+112%), accompanied by an almost double increase in volume (+98%), while volatility increased more moderately (+15%). The narrative built up in the preceding weeks and the strong anticipation for Musk's participation in SNL generated a long wave of speculation. However, immediately after the broadcast, which turned out pretty dry, the price underwent a rapid correction (-34%) and volumes and volatility also fell, confirming the presence of a classic ‘buy the rumour, sell the news’ phenomenon.

The data clearly show that the utterances of finfluencers, especially those with a large and active ‘fanbase’ like Musk, have the ability to significantly **alter** price, liquidity and market volatility in the short term.

In any case, this brief analysis has several limitations. Unlike traditional equities, the cryptocurrency market is inherently extremely volatile and sensitive even in the absence of finfluencer events. In addition, the temporal correlation between tweets and market movements is evident, but a linear cause-and-effect relationship cannot be proven with certainty. As shown by the first tweet, mass movements triggered by the Reddit communities were already taking place during the period analysed, so it is not possible to isolate the effect of Musk's posts.

In any case, even considering the example given, it is more than enough to make the reader aware of the enormous power of finfluencers over market dynamics and thus the need, as mentioned in previous chapters, for consistent and stringent regulation to prevent possible illicit behaviour and fraud.

4. FINAL CONSIDERATIONS

This thesis set out to investigate the ways in which social networks and digital influencers are reshaping financial markets, focusing on their effects on market volatility, liquidity and investor behaviour. Through a quantitative analysis of GameStop's (GME) short squeeze and a focused case study on the influence of a major 'finfluencer' in the cryptocurrency space, the research provided new empirical support for the thesis that online communities and digital personalities have become powerful agents in modern finance.

Evidence shows that collective sentiment on social media platforms such as Reddit can be **correlated** with extreme fluctuations in both trading activity and market instability. The GME episode in particular highlighted how coordinated, emotionally charged online discussion can lead to a rapid surge in trading volume and liquidity, posing a direct challenge to standard models of market efficiency and rational behaviour. Likewise, the Dogecoin paper demonstrated that a single powerful individual was able to use social media to induce extreme and rapid movements in price, trading volume, and volatility, which frequently deviate from traditional metrics or underlying value.

However, this work also highlights the methodological and interpretative challenges that arise from analysing digital sentiment and its impact on the market. It is necessary to recognise issues such as data quality, the limitations of sentiment analysis tools (especially in the context of informal, meme-rich language) and the highly endogenous nature of financial markets. In the case of cryptocurrencies, for example, the inherently volatile and speculative nature of the market can amplify or obscure the effects of digital influence, complicating causal inference.

Despite these limitations, the results demonstrate not only the need for further interdisciplinary research at the intersection of behavioural finance, data science and digital sociology, but also the urgent need for **regulatory frameworks** capable of addressing the new risks posed by online herd behaviour and finfluencer-driven speculation.

More broadly, this thesis feeds into larger discussions about the nature of financial markets in the digital era and the challenges of a new behavioural finance. It celebrates the fact that investor behaviour, the stability of markets and the determination of prices can no longer be explained only in terms of traditional financial theory but need to be examined in light of the emotional, controversial and network-related dynamics that constitute today's financial reality.

Future research and policy should continue to build on these insights, seeking to better understand and, where necessary, mitigate the transformative impact of social networks and finfluencers on global markets.

APPENDIX

A.1 Dataset

The Reddit WallStreetBets Comments dataset, created by Mateusz Podolak and made publicly available on Kaggle in 2021, collects and organizes comments from the popular subreddit r/wallstreetbets over an extended period. Podolak collected the comments using the Pushshift Multithread API Wrapper (PMAW), a well-known resource for accessing historical Reddit data. The extraction process involved querying the API for comments posted on r/wallstreetbets up to Feb 16, 2021. The dataset includes both the content of each comment and a set of associated metadata. It contains over 33 million unique comments described by 40 parameters. Here are some examples of the data contained in each comment:

- **id**: unique identifier for each comment.
- **author**: Reddit username of the comment author (anonymised).
- **body**: textual content of the comment.
- **created_utc**: date and time of comment publication (in UTC).
- **score**: number of positive votes minus negative votes (i.e. net score) received by the comment.
- **permalink**: direct link to the comment on Reddit.
- **parent_id**: identifier of the main comment or post (to reconstruct discussion threads).
- **link_id**: identifier of the associated Reddit contribution (post).

A.2 Python and libraries

Python was chosen for this analysis due to its clarity, versatility, and the wide range of powerful tools available for data science. Its simple syntax and large community make it easy to write, understand, and troubleshoot code, while its ability to integrate with different data sources and platforms ensures flexibility throughout the analytical process. Instead of writing every function from scratch, Python users can rely on libraries, or collections of predefined code, to efficiently perform specialised tasks such as data manipulation or numerical computation. Here is a complete list of the main Python libraries used during research and data analysis:

- `pandas` (data manipulation and analysis)
- `numpy` (numerical operations)

- `matplotlib` and `seaborn` (data visualisation)
- `yfinance` (market data retrieval from Yahoo! Finance)
- `vaderSentiment` (sentiment analysis)
- `emoji` (emoji conversion and mapping)
- `re` (regular expressions)
- `scipy` (statistical analysis)
- `datetime` (date handling)

A.3 Emoji handling

Instead of ignoring emojis or keeping them in their original state (their raw Unicode form), where most text analysis systems might not be able to understand their meaning, all emojis in Reddit comments were initially converted into standard text labels, such as ":rocket:", ":diamond_hands:" or ":fire:". This was made possible by using the specialised `emoji` library, that systematically translates each symbol into a standard text code. These codes were then converted into short descriptive words that more accurately describe what the emoji represents in the WallStreetBets group. For example, ":chart_increasing:" was replaced by 'profit gain rise', clearly referring to the growth of a stock, or ":eyes:" was transformed into 'watching attention', indicating a possible state of alert. This two-step process served to ensure that the unique slang and emotional content of the WSB community were accurately represented in the sentiment analysis. In this way, the analysis was able to detect not only common words, but also the playful, meme-based language that is critical to assessing the true sentiment of the community. This approach increases the accuracy and cultural appropriateness of the results, as it takes into account the specific way in which the WallStreetBets users express optimism, risk appetite or enthusiasm using emojis. Below is the list of the mapping used to convert emojis:

```
":rocket:": "rocket to the moon",
":fire:": "hot hype",
":clapping_hands:": "applause respect",
":thumbs_up:": "good approve",
":thumbs_down:": "bad disapprove",
":money_bag:": "money gain rich",
":chart_increasing:": "profit gain rise",
":chart_decreasing:": "loss crash down",
":pile_of_poo:": "shit bad horrible",
":gem_stone:": "diamond hands hold",
":hundred_points:": "100 perfect strong",
":screaming_face:": "panic shocked",
":grinning_face:": "happy smile",
":face_with_tears_of_joy:": "funny happy",
```

```
    ":red_heart:": "love like heart",
    ":broken_heart:": "sad loss heartbreak",
    ":skull:": "dead crash loss",
    ":face_with_symbols_on_mouth:": "angry mad",
    ":party_popper:": "celebrate happy",
    ":man_facepalming:": "frustrated regret",
    ":woman_facepalming:": "frustrated regret",
    ":face_vomiting:": "disgust awful",
    ":crying_face:": "cry sad loss",
    ":loudly_crying_face:": "cry hard loss devastated",
    ":thinking_face:": "thinking uncertain",
    ":eyes:": "watching attention",
    ":check_mark_button:": "agree yes",
    ":cross_mark:": "no disagree",
    ":clown_face:": "clown dumb joke",
    ":brain:": "smart big brain",
    ":face_with_monocle:": "analyze evaluate",
    ":folded_hands:": "hope pray",
    ":nail_polish:": "cool classy",
    ":sleeping_face:": "boring sleepy",
    ":smiling_face_with_sunglasses:": "cool confident",
    ":exploding_head:": "mind blown shocked",
    ":anguished_face:": "worried anxious",
    ":face_with_rolling_eyes:": "annoyed sarcastic",
    ":star_struck:": "amazed excited",
    ":partying_face:": "party success celebration",
    ":zany_face:": "crazy wild",
    ":face_with_head_bandage:": "hurt loss",
    ":nerd_face:": "smart analysis",
    ":face_screaming_in_fear:": "scared panic",
    ":face_with_thermometer:": "sick loss",
    ":trophy:": "win victory",
    ":alarm_clock:": "alert urgency",
    ":money_with_wings:": "loss money flying",
    ":face_with_cowboy_hat:": "yee haw ride",
    ":broom:": "sweep clear",
    ":sparkles:": "exciting hype",
    ":bellhop_bell:": "alert news",
    ":robot:": "bot automatic fake",
    ":hourglass_not_done:": "waiting delay"
```

A.4 VADER for sentiment analysis

For sentiment analysis in this study, the VADER (Valence Aware Dictionary and sEntiment Reasoner) algorithm was chosen as the primary tool. This tool is particularly well suited to analysing social media content and other short, informal texts, exactly the type of language found on platforms such as Reddit. Unlike classical sentiment analysis methods, which often require extensive training data and complex pre-processing, VADER is a rule-based model built on a judiciously curated dictionary of words, slang, emojis (although, as mentioned above, the emojis in our dataset were converted for greater accuracy) and even punctuation patterns, each associated with a specific sentiment score.

When VADER processes a comment, it scans the text for words and expressions known in its lexicon and assigns each of them a score that reflects its emotional tone, ranging from negative to positive. The algorithm then combines these scores, taking into account not only the words themselves, but also contextual clues such as punctuation (e.g., exclamation marks for emphasis), degree modifiers (e.g., “very good” vs. “good”) and negations (e.g., “not bad”). VADER is also designed to handle the informal language, abbreviations, and emotional expressions typical of online conversations, making it particularly effective for WallStreetBets comments. For each comment, VADER produces four separate sentiment scores:

- **Negative** (neg): the proportion of the text that conveys negative sentiment,
- **Neutral** (neu): the proportion that is neutral,
- **Positive** (pos): the proportion that is positive,
- **Compound** (com): an overall composite score that summarises the entire emotional tone of the comment, normalised between -1 (most negative) and +1 (most positive).

The negative, neutral, and positive scores each indicate how much of the text falls into those respective categories and always sum to 1. The compound score is a weighted aggregate that takes all the detected sentiment signals into account, providing a single, easy-to-interpret number for the overall sentiment. The combination of speed, ease of use and excellent performance on social media data makes VADER the perfect choice for this analysis, ensuring that both explicit statements and the unique communication style of the Reddit community are accurately interpreted.

For example, the comment: *"I'm not selling GME, I truly believe it's going to recover soon!"*

Returns, if analysed by VADER:

```
{'neg': 0.0, 'neu': 0.775, 'pos': 0.225, 'compound': 0.4926}
```

This shows how VADER is able to capture both explicit and subtle expressions of sentiment in social media texts.

A.5 The choice of parameter k for the weighting function

The parameter k in the weighting function acts as a decay factor that determines how quickly the influence of a comment decreases as its net score becomes more negative. In other words, k controls how strongly the weight assigned to each comment decreases for comments with low scores or strongly negative votes. Note that k only affects negative comments, because the logarithmic function does not depend on it. Setting a lower value for k causes the weight to decrease more rapidly with each negative vote, while a higher value would make the decay more gradual.

For this analysis, the parameter was set to $k = 10$. This specific choice was made after practical testing, as it offered a balanced compromise: it ensures that comments with many negative votes, those likely not in line with the community sentiment, have a minimal impact on the aggregate sentiment, while comments with only slightly negative or positive scores are still counted significantly, preserving the diversity of opinions. For example, with $k = 10$, a comment with a net score of -10 would receive a weight of 0.37, while a comment with a net score of -100 is almost completely ignored (~0.0000372).

If a moderately different value were chosen, such as $k = 9$ or $k = 11$, the results would not change dramatically: the penalty rate for negative scores would only be slightly higher (for $k = 9$) or more gradual (for $k = 11$), but the overall effect remains consistent.

Here are some practical examples of the weights assigned per score using the function:

Table 5 - Weight per given score

Net Score (s)	Weight
1000	7.91
100	5.61
10	3.30
0	1.00
-1	0.90
-10	0.37
-50	0.0067
-100	~0.0000372

REFERENCES

Alpert, M., & Raiffa, H. (1982*). A progress report on the training of probability assessors*. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under Uncertainty: Heuristics and Biases (pp. 294–305). Cambridge: Cambridge University Press.

Antweiler, W., & Frank, M. Z. (2004). *Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards.* The Journal of Finance, 59(3), 1259–1294

Apesteguia, J., Oechssler, J., & Weidenholzer, S. (2019). *Copy trading*. Management Science 66(12):5608-5622.

Ariely, D., Loewenstein, G., & Prelec, D. (2003). *Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences*. The Quarterly Journal of Economics, 118(1), 73–105.

Ashraf, N., Camerer, C. F., & Loewenstein, G. (2005). *Adam Smith, Behavioral Economist*. Journal of Economic Perspectives, 19(3), 131–145.

Baker, M., & Wurgler, J. (2007). *Investor sentiment in the stock market*. Journal of Economic Perspectives, 21(2), 129–152.

Barber, B. M., & Odean, T. (2001). *Boys will be boys: Gender, overconfidence, and common stock investment.* The Quarterly Journal of Economics, 116(1), 261–292.

Barber, B. M., & Odean, T. (2008). *All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors*. Review of Financial Studies, 21(2), 785–818.

Barberis, N., & Thaler, R. (2003). *A Survey of Behavioral Finance*. Handbook of the Economics of Finance: Volume 1B – Financial Markets and Asset Pricing, 1053–1128.

Berger, J., & Milkman, K. L. (2012). *What Makes Online Content Go Viral?* Journal of Marketing Research, 49(2), 192–205.

Berger, J., & Milkman, K. L. (2013). *Emotion and Virality: What Makes Online Content Go Viral?* GfK Marketing Intelligence Review, 5(1), 18–23.

Bernales, A., Valenzuela, M., & Zer, I. (2023*). Effects of Information Overload on Financial Markets: How Much Is Too Much?* International Finance Discussion Papers, No. 1372. Board of Governors of the Federal Reserve System.

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.

Bodie, Z., Kane, A., & Marcus, A. J. (2021). *Essentials of Investments* (12th ed., Chapter 9). McGraw-Hill Education.

Bollen, J., Mao, H., & Zeng, X. (2011). *Twitter mood predicts the stock market*. Journal of Computational Science, 2(1), 1–8.

CFA Institute & FINRA Investor Education Foundation. (2023). *Gen Z and Investing: Social Media, Crypto, FOMO, and Family.*

Cooley, D., & Parks-Yancy, R. (2019) *The Effect of Social Media on Perceived Information Credibility and Decision Making.* Journal of Internet Commerce, 18(3), 249-269.

Curme, C., Preis, T., Stanley, H. E., & Moat, H. S. (2014). *Quantifying the semantics of search behavior before stock market moves*. Proceedings of the National Academy of Sciences, 111(32), 11600–11605.

Cyert, R. M., & March, J. G. (1963). *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall.

Da, Z., Engelberg, J., & Gao, P. (2015). *The sum of all FEARS: Investor sentiment and asset prices.* Review of Financial Studies, 28(1), 1–32.

De Bondt, W. F. M., & Thaler, R. (1985). *Does the stock market overreact?* The Journal of Finance, 40(3), 793–805.

De Bondt, W. F. M., & Thaler, R. (1987). *Further evidence on investor overreaction and stock market seasonality.* The Journal of Finance, 42(3), 557–581.

Edwards, W. (1968). *Conservatism in human information processing*. In Kleinmutz, B. (Ed.), Formal Representation of Human Judgment . New York: Wiley, 17-52.

Espeute, S., & Preece, R. (2024). *The Finfluencer Appeal: Investing in the Age of Social Media*. CFA Institute: Industry Future Series.

FCA (2021). *Understanding self-directed investors: Research and findings.* Financial Conduct Authority.

Fama, E. F. (1970). *Efficient Capital Markets: A Review of Theory and Empirical Work*. Journal of Finance, 25(2), 383–417.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). *Knowing with certainty: The appropriateness of extreme confidence*. Journal of Experimental Psychology: Human Perception and Performance, 3(4), 552–564.

Gervais, S., Kaniel, R., & Mingelgrin, D. (2001). *The high-volume return premium*. Journal of Finance, 56(3), 877–919.

Gilovich, T., Vallone, R., & Tversky, A. (1985). *The hot hand in basketball: On the misperception of random sequences*. Cognitive Psychology, 17(3), 295–314.

Hayes, A. S., & Ben-Shmuel, A. T. (2024). *Under the finfluence: Financial influencers, economic meaning-making and the financialization of digital life*. Economy and Society, 53(3), 478–503.

Hou, K., Peng, L., & Xiong, W. (2009). *A tale of two anomalies: The implications of investor attention for price and earnings momentum.*

Ivantchev, B., & Ivantcheva, M. (2023). *FOMO effect: social media and online traders*. Journal of Management and Financial Sciences, (48), 59–74.

Kahneman, D., & Tversky, A. (1974). *Judgment under Uncertainty: Heuristics and Biases*. Science, 185(4157), 1124–1131.

Kahneman, D., & Tversky, A. (1979). *Prospect Theory: An Analysis of Decision under Risk*. Econometrica, 47(2), 263–291.

Kahneman, D., & Tversky, A. (1982). *The Psychology of Preferences*. Scientific American, 246(1), 160–173.

Kahneman, D., & Tversky, A. (1984). *Choices, values, and frames*. American Psychologist, 39(4), 341–350.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money*. London: Macmillan, Ch. 12.

Khashanah, K., & Alsulaiman, T. (2016). *Network Theory and Behavioral Finance in a Heterogeneous Market Environment.* Stevens Institute of Technology, Financial Engineering.

Kirman, A. (1993). *Ants, rationality, and recruitment*. The Quarterly Journal of Economics, 108(1), 137–156.

Kumar, S., Roy, P. P., Dogra, D. P., & Kim, B.-G. (2023). *A comprehensive review on sentiment analysis: Tasks, approaches and applications* Journal of Economic Surveys, 37(2), 444–478.

Lev, B., & Gu, F. (2016). *The End of Accounting and the Path Forward for Investors and Managers* (1st ed., Chapter 8). Wiley.

Malmendier, U., & Tate, G. (2008). *Who makes acquisitions? CEO overconfidence and the market's reaction.* Journal of Financial Economics, 89(1), 20–43.

Oldham, M. (2019). *Understanding how short-termism and a dynamic investor network affects investor returns: An agent-based perspective*. Complexity, 2019

Oosting, J. (2022). *Finfluencers and their impact on the stock market* (Bachelor's thesis, Leiden Institute of Advanced Computer Science, LIACS, Leiden University).

Plous, S. (1993). *The Psychology of Judgment and Decision Making*. New York: McGraw-Hill.

Podolak, M. (2021). *Reddit WallStreetBets Comments* [Data set]. Kaggle. https://www.kaggle.com/datasets/mattpodolak/reddit-wallstreetbets-comments

Preis, T., Moat, H. S., & Stanley, H. E. (2013). *Quantifying Trading Behavior in Financial Markets Using Google Trends*. Scientific Reports, 3, 1684.

Scharfstein, D. S., & Stein, J. C. (1990). *Herd behavior and investme*nt. The American Economic Review, 80(3), 465–479.

Selden, G. C. (1912). *Psychology of the Stock Market: Human Impulses Lead to Speculative Disasters.* New York: Ticker Publishing.

Sharpe, W.F. (1964) *Capital Asset Prices: a Theory of Market Equilibrium Under Conditions of Risk*. Journal of Finance 19(3), 425-442.

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.

Shiller, R. J. (2003). *From Efficient Markets Theory to Behavioral Finance*. Journal of Economic Perspectives, 17(1), 83–104.

Shiller, R. J. (2019). *Narrative economics: How stories go viral and drive major economic events.* Princeton University Press.

Shiller, R. J., & Pound, J. (1989). *Survey Evidence on Diffusion of Interest and Information Among Investors.* Journal of Economic Behavior and Organization, 12, 47–66.

Shive, S. (2010). *An Epidemic Model of Investor Behavior*. Journal of Financial and Quantitative Analysis, 45(1), 169–198.

Shleifer, A., & Vishny, R. W. (1997). *The Limits of Arbitrage.* The Journal of Finance, 52(1), 35–55.

Simon, H. A. (1955). *A Behavioral Model of Rational Choice*. The Quarterly Journal of Economics, 69(1), 99–118.

Simon, H. A. (1978). *Rational Decision-Making in Business Organizations*. Nobel Prize Lecture.

Simon, Herbert A. (1971). *Designing Organizations for an Information-rich World*. In M. Greenberger (Ed.), Computers, Communications, and the Public Interest, Baltimore, MD: Johns Hopkins University Press, 37-52.

Singer, E. (2002). *The use of incentives to reduce nonresponse in household surveys*. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), Survey Nonresponse, Wiley, 163–177.

Singh, A., & Sarva, R. (2024). *The rise of finfluencers: Financial influencers and the digital investor journey.* Australasian Accounting, Business and Finance Journal, 18(3), 268–278.

Sloan, R. G. (1996). *Do stock prices fully reflect information in accruals and cash flows about future earnings?* The Accounting Review, 71(3), 289–315.

Smith, A. (1759). *The Theory of Moral Sentiments.*

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). *Tweets and Trades: The Information Content of Stock Microblogs.* European Financial Management, 20(5), 926–957

Săvoiu, G., & Iorga Simăn, I. (2013). Sociophysics: *A new science or a new domain for physicists in a modern university.* In G. Săvoiu (Ed.), Econophysics, San Diego, CA: Academic Press, 149–168.

Tetlock, P. C. (2007). *Giving Content to Investor Sentiment: The Role of Media in the Stock Market.* The Journal of Finance, 62(3), 1139–1168

Thaler, R. H., & Johnson, E. J. (1990). *Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice*. Management Science, 36(6), 643–660.

Tsintsaris, D.; Tsompanoglou, M.; Ioannidis, E.(2024) *Dynamics of Social Influence and Knowledge in Networks: Sociophysics Models and Applications in Social Trading, Behavioral Finance and Business.* Mathematics , 12, 1141

Tversky, A., & Kahneman, D. (1981). *Judgments of and by representativeness* (Report). Stanford University.

Weinstein, N. D. (1980). *Unrealistic optimism about future life events*. Journal of Personality and Social Psychology, 39(5), 806–820.