

TRADING GAMIFICATION

BENOIT BORDACHAR

18 December 2021

CONTENTS

1	Introduction	3
2	The Gamification Phenomenon	4
2.1	Theory	4
2.2	Application to stock trading	5
3	Manifestation of Trading Gamification	9
3.1	Origin	9
3.2	Characteristics	11
3.3	Empirical Results	12
4	Use of Trading Gamification	13
4.1	Sentiment Analysis	14
4.2	Classification Model	18
5	Conclusion	20

LIST OF FIGURES

Figure 1	Robinhood lottery	5
Figure 2	The Octalysis framework	6
Figure 3	Retail investors trading volume	10
Figure 4	Retail investors net purchases	10
Figure 5	Wordcloud of Jan 2021 WSB comments	14
Figure 6	Sentiment signal	16
Figure 7	Fourier transform of the sentiment	16
Figure 8	Rolling correlation sentiment signal and SPY	17
Figure 9	Models architecture	19

LIST OF TABLES

Table 1	Table of Correlations	17
---------	---------------------------------	----

ABSTRACT

After a year marked by the resurgence of retail traders and the saga of meme stocks such as GAMESTOP which left many institutional investors puzzled, I wanted to understand one of their key-drivers: the gamification of trading. Rather than weighing up the pros and cons of its existence, I chose to take it as a fact and find its empirical manifestations in order to take it into account. To do so, I used the comments of the subreddit *wallstreetbets* through various analysis and natural language processing models. In the end, it appears that there is a relationship to be exploited between the SP500 and WSB, however it would not be sufficient on its own. Ultimately, this work argues for the inclusion and modelling of trading gamification to approach the current financial markets.

1 INTRODUCTION

Earlier this year, ROBINHOOD ditched its confetti animation¹ after accusations of being reckless toward its newbie investors by making light of decisions that involve real money.

Last August, the Securities and Exchange Commission deepened its investigation into purportedly *gamification* and behavioral prompts used by online brokerages such as ROBINHOOD mobile application to spur people to trade more actively stocks and other securities.

SEC² Chair G. Gensler said in a release [13]: "*While new technologies can bring us greater access and product choice, they also raise questions as to whether we as investors are appropriately protected when we trade and get financial advice, ... In many cases, these features may encourage investors to trade more often, invest in different products, or change their investment strategy.*". The SEC is curious about the need for the financial regulator to protect retail investors from a potential conflict of interest with online brokers. Indeed, online brokers are remunerated with the commissions taken on the trades made on the platform. Thus, for the SEC, the question is whether incentivizing unwary investors with game-like, or predictive prompts assume optimal outcomes, to place more orders would generate a loss in investors' portfolio performance.

This SEC survey demonstrates the magnitude of the gamification of trading. Between the development of more user-friendly applications with the aim of democratising finance and manipulating a young, uninformed public, the SEC will have to establish a regulatory framework for online brokers and the use of gaming codes.

Here, I will not seek to know if brokers are at fault and measures should be applied, but rather to understand the ins and outs of this phenomenon in order to build a trading strategy on it. Some argue that the gamification of trading only add an irrational factor to the stock market by oversimplifying stock trading, and that this young public should be educated to avoid the bursting of an irrationality bubble.

Whether one is for or against the gamification of trading, it is here and now and must be factored into the understanding of the market. The new type of investor drawn to gamification, mainly young

¹ ROBINHOOD used to shower digital confetti down a smartphone screen upon successful execution of a trade

² Securities and Exchange Commission

and lacking in experience, has the capacity to stir up the markets, as witnessed in the GAMESTOP case.

In this paper I will first try to explain the phenomenon of gamification of trading and then study its characteristics and empirical results in the markets. Finally, I will propose a way to implement gamification of trading to account for it in the approach to financial markets.

2 THE GAMIFICATION PHENOMENON

2.1 Theory

DEFINITION Although there is no unanimous definition yet, gamification is defined as “the use of game mechanics and experience design to digitally engage and motivate people to achieve their goals” [2]. In other words, gamification is a behavioral design that uses familiar feature of the online world. These game design elements are applied to motivate and engage individuals in non-game contexts. Moreover, gamification refers to game elements that do not constitute a whole game. In practice, the design of a gamified application will most likely give rise to playful mindsets and behaviors (Deterding et al., 2011, [5]). For example, leaderboards, badges, prices and animations are considered to be game design elements.

CONSEQUENCE The possible effects of gamification have been pinpointed by Andrade and others (2016) [1]. One of them is the *off-task behavior*, as they call it, which consists of game elements not linked to the platform’s purpose that drag off users’ attention. For instance, some customization features may encourage the user to spend time on the platform, but may not help them to achieve the platform’s objective.

The gamification might also create *undesired competition*, especially when using leader-boards users feel forced to compete against other peers.

Above all, there is also the *addiction and dependence* engendered. For example, the user may be more focused on collecting more points or climbing the leader-boards rather than focusing on the real purpose of the gamified system. The user could also become dependent on extrinsic rewards in order to stay engaged. Furthermore, a sensation like flow created by immersive app-design, can be addictive just as various game features.

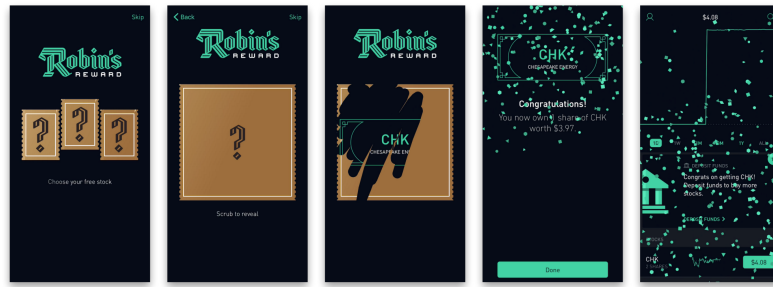


Figure 1: Illustration of ROBINHOOD rewards

At the end, a user of gamified platform may develop an addictive consumption and have a disproportionate uncontrollable desire to make purchases. Addictive consumption especially arises as a way to avoid or overcome various feelings such as anxiety, depression or tension.

2.2 Application to stock trading

2.2.1 Use of gaming mechanics

The idea behind gamification of stock trading is to take advantage of the techniques used to make games successful and apply them to trading platforms to make them more fun and engaging. By doing so, online brokers hope to boost the volume of transactions processed on their platform and generate more commissions. Indeed, the more fun and engaged a person is when using the application, the more motivated they are to use it.

Trading applications try to have an aesthetically pleasing or "eye candy" design. ROBINHOOD's signature piece of "eye candy" was the digital confetti which has now been removed. Upon completion of a trade, confetti used to rain down the screen (see Figure 1)

Online brokers also find themselves offerings rewards to their users. For example, ROBINHOOD gifts its users lotteries for potentially valuable surprise stocks as rewards for linking their bank accounts or referring new users. Besides lotteries, trading apps also reward their frequent engagers with privileged access to freshly released features.

Figure 1 shows the flow of screens that a user would experience during a lottery : a three cards monte, a scratch ticket, and a flurry of confetti. Visuals that evoke lottery-like phenomena, encouraging people to equate trading as having the potential to be "jackpot".

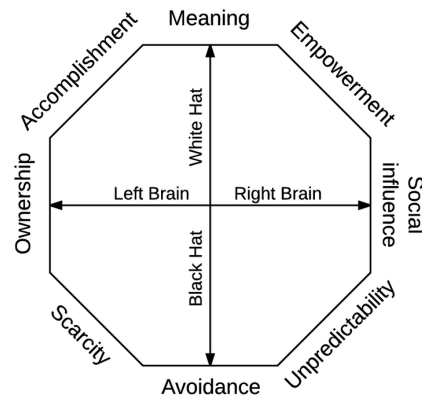


Figure 2: Illustration of the Octalysis framework by Chou (2016)

2.2.2 Experience design to influence behavior

To understand the gamification of stock trading, the *Octalysis* framework proposed by Chou (2016) [3] could be applied, it suggests there are eight core drives that motivate people to perform an action (Figure 2). The design of trading applications should answer these drivers in order to be "gamified".

The Octalysis can be split into 4 sub-parts :

- The left side is related to ownership, analytical thought, and logic. The individuals placed there tend to be motivated by extrinsic elements.
- The right side is associated with intrinsic motivation. Activities such as creativity, social-dynamics, and self-expression are rewarding by their own for individuals.
- The top core drives are considered *white hat* as they result in positive emotions and make individuals feel fulfilled, powerful, and satisfied.
- While the bottom core drives are *black hat* because these produce feelings of fear, surprise and disappointment. They make individuals feel obsessed, addicted, and anxious.

EPIC MEANING AND CALLING Meaning is something everyone looks for more of and awe is a product when one recognizes he or she is a part of something bigger than themselves [4]. For example, the famous r/wallstreetbets REDDIT blog, where users freely share their knowledge, their moves and trade ideas, to a larger community of users which leads to a sense of self-worth.

DEVELOPMENT AND ACCOMPLISHMENT Accomplishment is one of the most basic emotions one can feel. Usually, Accomplishment is what you feel after overcoming a challenge. It is this search for fulfilment that drives the development of one's skills. The more difficult the task, the more intense the feeling. For instance, ROBINHOOD deploys a user experience rooted in positive reinforcement structure. When account holders make their first deposit, they receive a congratulatory message saying that funds have been made available immediately so they can start trading.

EMPOWERMENT OF CREATIVITY AND FEEDBACK This driver aims to stimulate the creativity of the user by engaging him in a repetitive process of problem solving through trial and error. This is also true for portfolio construction, where the problem is to find the right combination of stocks to make a profit. By offering its users the possibility of creating lists of stocks to watch, ROBINHOOD feeds this reflection on the composition of the portfolio to adopt to solve this problem.

OWNERSHIP AND POSSESSION This driver is activated by the desire for possession and the feeling of ownership. People take a sense of pride in what they own and innately tend to make it better than it is. Therefore, the personal value of the object immediately increases. Also in ROBINHOOD, each time you buy a share, a message saying "Congratulations! You now own X share(s) of XYZ worth \$XXX" pops up on the screen app and seems to activate this driver perfectly.

SOCIAL INFLUENCE AND RELATEDNESS This driver plays with the human desire to connect and compete with other people to make it an intrinsic motivation. In doing so, people may be tempted to modify their behaviour according to others. ETORO, another online broker, exploits the mass friend-spamming mechanics with community features that allows users to share their stats and make comments on each other's trading activity.

SCARCITY AND IMPATIENCE The very nature of the desire to possess is to crave objects that we don't have, and it's all the stronger when these objects are difficult to get because of unavailability or high cost. This driver pushes for great and urgent consumer action however this sometimes leads to irrational actions. When ROBINHOOD first launched, it played on the FOMO³ by making potential customer feel like they might be missing out an opportunity to get rich. ROBINHOOD harnessed FOMO by creating an impression of exclusivity through the use wait lists. In addition, prospective users

³ Fear Of Missing Out

could engage with “a referral-based viral loop” to climb up the wait-list by referring other prospective users.

UNPREDICTABILITY AND CURIOSITY Curiosity is the main impulse and driving force behind people’s infatuation with uncertain experiences. Unpredictability adds a measure of anticipation and there is an emotional build-up when waiting for something unpredictable to occur. The brain becomes so engaged in this process, that the person hooked on this unpredictable chance to win. To arouse this curiosity, trading platforms did not need much more than to display price graphs in the most eye-catching way possible as curiosity rushes interest when exposed to a visual stimulus. After all, it is the essence of trading to attempt to anticipate what a stock is going to do.

LOSS AND AVOIDANCE This core driver motivates people by inducing the fear of losing possession of something or the fear of undesirable things happening. It can be exploited to guide users towards performing desired actions. For example, stock trading platforms send push notifications to signal sharp declines or gains, or events in the market, to encourage the user to place orders.

Through the lens of Octalysis framework, it appears that online brokers have gamified their interface in the purpose of influencing users. Because of the stakes involved in trading and the drivers that trading platforms push, users of these applications would tend to be in the bottom side of the octagon.

The design of the trading interface combined with the display of specific messages seems to be made in such a way as to reinforce the drivers of scarcity, avoidance and unpredictability. These core drives are considered as negative motivations because of the obsession, addiction and anxiety they spawn. Here, the meaning driver rather activated on social networks like TWITTER and especially REDDIT allows investors to stay motivated to trade despite the bad taste left by the black hat drivers. *“For those making large bank on this I congratulate you. For those who may lose a little, F it YOLO⁴ and this is epic history in the making. Go get em!!!!”* should we read on r/wallstreetbets in REDDIT. In addition, there is the social driver which supports the investor in his quest for meaning, even if succumbs to the temptation of a herd-like behaviour and a crowd mentality.

The stock market, already accused of being a vast casino with many inherent psychological issues, needed only a few codes from the world of video games to finally take on its full features. In the end, trading platforms seem to play on the psychological challenges

4 “you only live once”

of trading, like fear and greed, which are already present but not fully visible, to better strengthen them through a gamification of their interface.

3 MANIFESTATION OF TRADING GAMIFICATION

As it has been introduced, zero-commission trading platform such as ROBINHOOD innovated in user-experience design, featuring game practices — flashy graphics, rewards, and the like — that make it attractive, easy, and fun to trade stocks.

After tracing its origin, let's try to understand the underlying characteristics of gamification of trading and its empirical results in the financial markets.

3.1 Origin

In my opinion, trading gamification can be traced back to 2019 when online brokerages eliminated fees in a movement of digitalization and decentralization of asset management. However, it truly took off with the COVID-19 pandemic. The gamification of trading has been such that the literal meaning of the term is now experienced. *Metaphorical gaming* as J. F. Tierney calls it [9]. It appears that *"easy-and-free trading, combined with expressive aspects of coordinating with other traders, makes a 'game' of trading"*. In this way, the notion of meme stocks emerged in 2020 as a consequence of the gamification of investing. Social media like REDDIT and TWITTER have helped build hype around stocks through narratives or elaborated discussion threads, to such an extent that some shares of companies have gained a cult-like and are now referred as "meme stocks". These meme stocks are considered to be "in-app purchases" by Matt Levine, famous market commentator. In meme stock herding trades, like with other games, it is possible to pursue another expressive, performative and "gameful" end that doesn't necessarily involve making money-like.

The gamification of trading is for J.F Tierney the product of convergent trends : "re-retailization" of capital markets, price competition on brokerage fees and fragmentation of market structure.

RE-RETAILIZATION Recently, there has been a resurgence of interest in trading among retail investors, helped in particular by the various lockdown periods. In this view, trading might have served as a substitute for other kinds of entertainment. This re-emergence has resulted in an increase in the share of the traded volume (see Figure 3 on the next page and 4 on the following

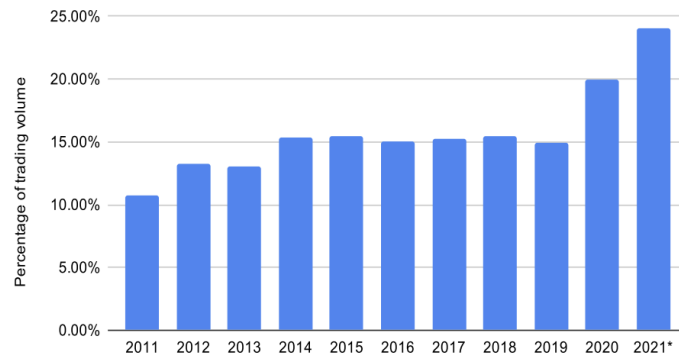


Figure 3: Retail investors' share of US equities trading volume

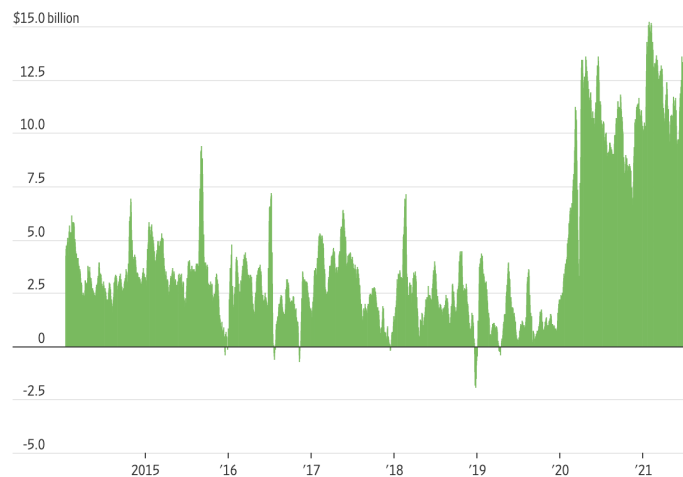


Figure 4: Retail investors' net purchases of US stocks and ETFs, 10-day sum

page). However, one interesting fact is that this increase in trading volume did not coincide with a corresponding jump in ownership share of total market value, showing the short-term vision of retail investors and their overtrading tendency.

BROKER COMPETITION Behavioral design of trading apps is also due to the competition between brokers on commissions. Commission pricing is a key factor in the decision making process to choose which brokerage services to opt for. Since many brokers cut commissions to zero in 2019, the only thing left to do was to fight over the design of the applications, especially as people like to focus on salient attributes. With the arrival of digitally savvy younger traders, brokers had every reason to incorporate those behavioral design in their user experience.

MARKET FRAGMENTATION Thanks to a fragmented market across dispersed and interlocking set of rules, institutions and practises, possibilities have arisen for sophisticated market participants to bridge gaps, providing liquidity while using information about retail order flow for profit. To afford zero-commission offer, bro-

kers had to find another source of revenue : PFOF⁵. This is the compensation for routing retail order flows to third-party principal trading firms for execution. By virtue of being "noisy", retail orders generates opportunities for these wholesale dealers who profit from arbitrage and taking the contra side. Although these payments need to be disclosed and must abide by the "best execution" brokers' duty, reality shows that sometimes brokers put their interests ahead of their clients'. For example, FINRA⁶ condemned ROBINHOOD for violation of the best execution principle [6].

3.2 Characteristics

Hence, the gamification of trading is characterised by the massive arrival of a rather young and uninformed public, with a certain appetite for easy short-term profits and risk with reckless behaviour, but which nevertheless has the capacity to move the markets. It is also important to take into account the possibility that this particular market movement could be pre-empted by institutional investors through the PFOF.

The traders brought in by gamification have a tendency to excessive trading generating noisy order flow, much to the blushing delight of brokers. An intriguing trait of these "gamer" traders is that even if they are losing, they might be trying to satisfy other non-pecuniary preferences. The first suggestion is that behavioral design encourages or tricks them into it. Still some do it for the entertainment or the thrill it gives them (as a gambler would do), while others aspire to riches through high volatility lotteries. There are those who trade because a stock is salient to them or because they have been lured into doing so. In general, the reasons for retail investors to trade do not always logically answer liquidity, tax or re-balancing needs. Even some gain emotional or affinity value from connecting online with like-minded traders. These investors seem to "play" the market filling non-pecuniary purposes, to the point of turning it into a game.

From these findings, one can infer that retail investors would be imperfectly rational or at least asymmetrically informed, thus contradicting the main assumptions of the efficient market. Their decision-making process would be subject to bounded rationality and will lead them to act noisily in ways uncorrelated with the market. For retail investors, engaging in fundamental analysis or research to learn private information that could be traded on for profit does not appear to

⁵ Payment For Order Flow

⁶ Financial Industry Regulatory Authority

be cost-effective. Retail -gamified- investors are prone to trade without superior information about a stock's fundamental value. This way, they prefer to operate as ordinary consumers however "noisy" their trading become with respect to payoff-relevant information. As a result, retail investors are very sensitive to the presentation of information, and gamification takes on its full relevance to brokers.

3.3 Empirical Results

The gamified trading practise is such that the market feels its presence as evidenced by the GAMESTOP saga : *"GameStop shares have soared 1,700 percent as millions of small investors, egged on by social media, employ a classic Wall Street tactic to put the squeeze — on Wall Street"* [12]. The interesting thing about gamification traders is that they behave as a group. And, even though losing traders are more likely to stop trading, empirical evidences have shown that losing traders nonetheless persist in the market as a group.

Philipp Chapkovski, Mariana Khapko and Marius Zoican showed that trading platform gamification increases retail investors' risk tolerance [11]. They found that the effect of trading gamification on risk tolerance was particularly strong with assets considered risky to begin with, such as options or crypto-currencies. Hence, trading platforms seems made to guide and tilt retail investors' attention towards high-volatility asset classes. The evidence is that online brokerage firms derive a large share of their profits from options rather than stocks. In addition, they noted a greater impact on inexperienced traders with a low level of financial knowledge, thus justifying the need for investor education and protection.

Gamified trading certainly produces noise as it may confuses or misrepresents genuine underlying trends (3.2). The short-term volatility created has the merit of providing liquidity to the market as some institutional firms can profit from trading against gamified traders. If we consider gamified trading to be noise and to be uncorrelated with news affecting the fundamentals of a stock, then it might promote price discovery. Theoretically this would attract more informed traders to restore balance between prices and fair values, but in practise Eaton, Green, Roseman, and Wu [7] this disrupts the price discovery process by increasing both price movement and volatility in stocks popular among retail investors.

Furthermore, brokers offer recommendations algorithms that refer a list of stocks to clients. This list may be constructed using criteria such as "top movers", high trading volume or most concentrated holdings among clients. As a result, these recommendations are inclined to

enhance salience of certain securities and induce demand by a phenomenon of *attention-induce noise trading*.

It can be argued that gamified trading, with the ease and encouragement of in-and-out trading through zero-commission brokerage, is detrimental to market quality in general and distort capital allocation in particular. Eaton, Green, Roseman, and Wu proved that “quoted spreads, effective spreads, realized spreads, and price impact” are higher quality when ROBINHOOD investors exit the market. Therefore, it might have downstream negative consequences on public confidence in financial markets.

4 USE OF TRADING GAMIFICATION

In the previous section, I attempted to exhibit how the gamification of trading has shaped a new kind of trader, say gamified traders. It is not uncommon for these traders to garner recognition as a group. The gamified traders also operate in a fashion that does not typically serve a pure profit motive, making trading a game in the eyes of some. A place where these traders like to meet to exchange and unite is */wallstreetbets* (WSB) a REDDIT sub-forum. Since the meme stocks saga exists, it has been occupying space in the mind of all investors and financial professionals. WSB gathers over 11,364,938 subscribers making the subreddit the 5th most visited website across the entire internet. The meme stocks proved gamified traders’ ability through WSB to move markets.

The assumption here is that there is a significant chance for WSB to generate herd mentality creating an effect on individual tickers and/or certain ETFs (Exchange Traded Funds). The goal of this section is to build a model in order to see if there is any predictive value within *r/wallstreetbets*. I think WSB is a good proxy for the trading gamification phenomenon. Although I must admit that I would have liked to include data from ROBINHOOD, but the online broker turned off the feature allowing anyone to see which companies’ shares were surging in popularity on the platform.

To gather the comments from WSB, I used a dataset collected from KAGGLE and I completed it with pushshift.io and PRAW (the REDDIT API). The database created includes all discussion threads with comments between 11 April 2012 and 22 December 2021. For you to have an idea about what look like the WSB comments, I join you a wordcloud of all the comments in January 2021, when the GAMESTOP case happened.



Figure 5: Wordcloud of Jan 2021 WSB comments

In a flurry of stock tickers, we can read the words "melvin" in reference to MELVIN Capital having suffered heavy losses after shorting the GAMESTOP share. There are also many other words related to the event, such as "deepfuckingvalue", a WSB account that is said to be the mastermind of the GAMESTOP operation, as well as "ladder" for the alleged short ladder attacks launched by institutional investors to counteract the meteoric rise of the stock.

To analyze REDDIT data, I will use NLP and data science techniques to check how much of interest WSB is. First I will explore the relationship between a sentiment signal from WSB comments and the SP500. This analysis will allow me to have a first insight on the value of WSB. Then I shall endeavour to build a classification model on the WSB comments for predicting the direction of the SP500 on the day after the daily discussion threads.

My first intention would have been to use the NASDAQ as its large share of Tech stocks was more likely to interest REDDIT users in my opinion, but the reality showed that the SP500 through the SPY⁷ was the most mentioned financial product on the forum (see <https://www.quiverquant.com/wallstreetbets/>).

4.1 Sentiment Analysis

In my opinion, the first thing interesting to do is to perform a sentiment analysis on the REDDIT comments and observe the correlation with the SPY across the given period. In my research I found an **article** of Arjun Rohlfing-Das that I think is interesting, so I took the liberty of partly following his method of analysis.

⁷ The SPY is an ETF designed to track the SP500

A sentiment analysis is a text analysis method that detects polarity (positive or negative opinion) within a text, here comments. It aims to measure the attitude, sentiments, evaluations, attitudes, and emotions of a speaker/writer based on the computational treatment of subjectivity in a text.

On each comment, I run the VADER algorithm (Valence Aware Dictionary for Sentiment Reasoning). VADER is a parsimonious rule-based model for sentiment analysis of social media text [8]. I have added words from the REDDIT vocabulary to increase the sentiment precision, even though VADER incorporates a "gold-standard" sentiment lexicon that is especially attuned to microblog-like contexts. The sentiment scores from all the comments of each day are then summed to create a daily sentiment score.

I focus my analysis on the period from 2019 to today, as I believe that 2019 really marks the dawn of the WSB forum. I did detect some outliers periods in the data, for example the GAMESTOP saga created a spike in the sentiment signal in January 2021.

To process these periods, I calculated the z-score of the signal series and capped its absolute value at 3. If the population mean and population standard deviation are known, a raw score x is converted into a standard score by :

$$Z = \frac{x - \mu}{\sigma}$$

where:

μ is the mean of the population,

σ is the standard deviation of the population.

On the Figure [6 on the next page](#)), I plotted the sentiment signal corrected from outliers, against the SPY prices history. We can see that the signal dropped with the SPY during the Covid-19 crash. However, from April 2021 onwards they seem to diverge.

Next, I applied discrete Fourier transforms to smooth the sentiment signal (See Figure [7 on the following page](#)). The Fourier Transform is a tool that breaks a waveform (a function or signal) into an alternate representation, characterized by the sine and cosine functions of varying frequencies. The Fourier transform will be far less noisy than the raw sentiment signal.

The new sentiment series is computed simply by using the following formula :

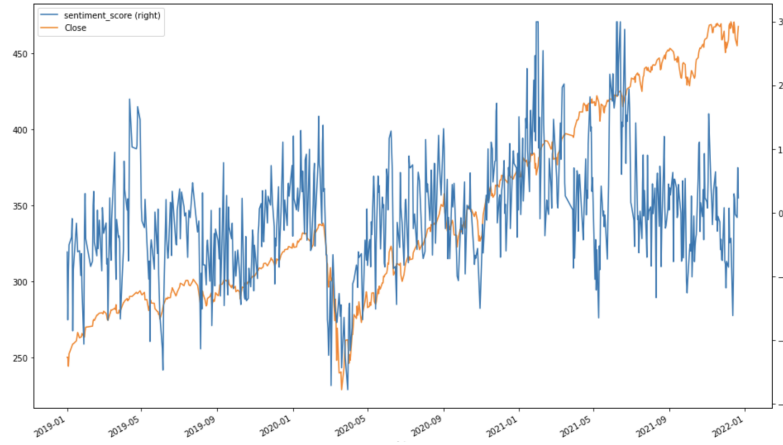


Figure 6: VADER sentiment signal of WSB comments

$$X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i k \frac{n}{N}}$$

As can be seen from the graph, this transformation allows us to see more clearly the relationship between the SPY and the sentiment signal.

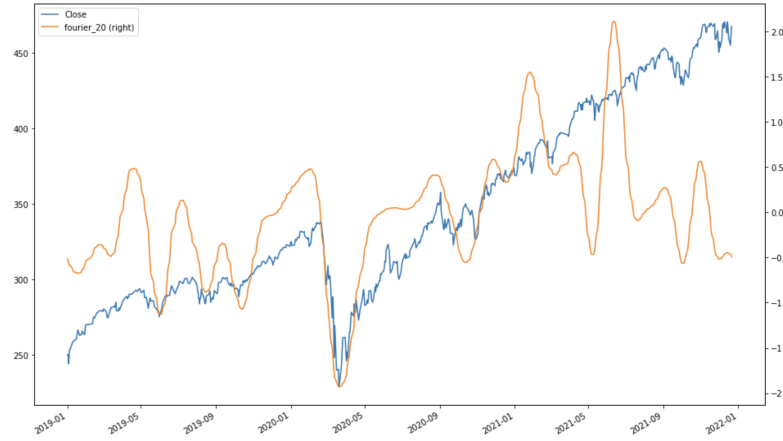


Figure 7: Fourier transform of the sentiment with 20 components

If we check the correlation between the normalized SPY prices and the normalized 20 components Fourier transform of the sentiment signal, we find that there is a correlation of 48%. The mean of the 14-days rolling correlation is 41% while the standard deviation is 55%. It seems that there is a good link between the SPY and the Fourier transformed sentiment signal, however the relation seems to variate a lot. Reference to Table 1 on the next page.

Table 1: Table of Correlations

Correlation Matrix		
	norm_spy_price	norm_fourier
norm_spy_price	1.0000	0.4757
norm_fourier	0.4757	1.0000

Rolling Correlation	
mean	0.410
st. dev	0.547

Lastly, I plot (see Figure 8) the rolling correlation over the normalized SPY price with red and black axis lines at the mean and mean minus standard deviation of the rolling correlation.

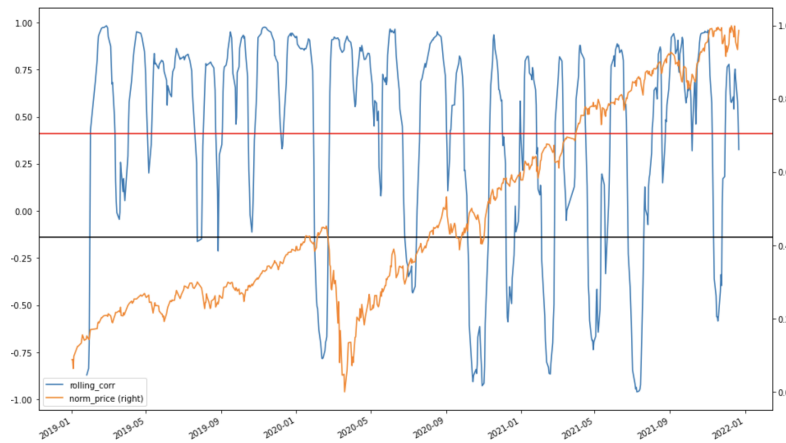


Figure 8: 14-d rolling correlation between SPY and sentiment signal

What's interesting is that a relationship seems to emerge from the SPY and the Fourier transform of the sentiment signal. Furthermore, the rolling correlation shows that the relationship is quite wavy, the SPY and the sentiment signal are either highly positively correlated or completely opposite. The relationship tends to be more positive than negative over time, resulting in a correlation of 48%.

Everything considered, the sentiment signal smoothed by a Fourier transformation alone will hardly be enough to predict the next-day direction of the SPY, but this analysis argues for considering the signal as a complementary factor in a broader model. Besides, it gives ground for implementing an NLP model on the WSB comments.

4.2 Classification Model

As I explained before, it might be worth exploring the forecasting power of the WSB comments on the SP500. So, I will try to build a model on the REDDIT comments to predict the next day SPY direction. I would like to know if the SPY is more likely to go up or down during the day following the WSB daily discussion thread.

This is a typical binary classification task in NLP, however the quality of the REDDIT data is certainly questionable and foretell the direction of the SP500 is a dream lot of people have.

As the database we have managed to assemble is relatively small, we will apply transfer learning, which is a technique where instead of training a model from scratch, we reuse a pre-trained model and then fine-tune it for another related task. Indeed, if we wanted to train a neural network model from scratch, we would likely over-fit the data. For this endeavour, we will rely on the [fast.ai](#) library through this method :



For both the language model and the classifier, [fast.ai](#) uses the AWD-LSTM architecture (ASGD Weight-Dropped LSTM) which is based on [DropConnect](#) and the average random gradient descent method, with several other regularization strategies (see [Figure 9 on the next page](#)).

Long Short-Term Memory (LSTM) is a kind of Recurrent Neural Network (RNN), which is suitable for processing and predicting sequential sequences, especially for NLP. LSTM cell structure consists of four component : forgotten gate, input gate, output gate and cell state. In the LSTM structure, σ represents a sigmoid function, \tanh represents a tanh function, $X(t)$ represents the input, $h(t-1)$ refers to hidden states, and $C(t)$ is the state of the cell. For more details, I refer you to this [paper](#).

Universal Language Model Fine-tuning for Text Classification (ULM-FiT [\[10\]](#)) is the pre-trained model. It relies on Wikitext which consists of a preprocessed subset of 103 million tokens extracted from Wikipedia. It's a model that understands a lot about language and a lot about what language describes.

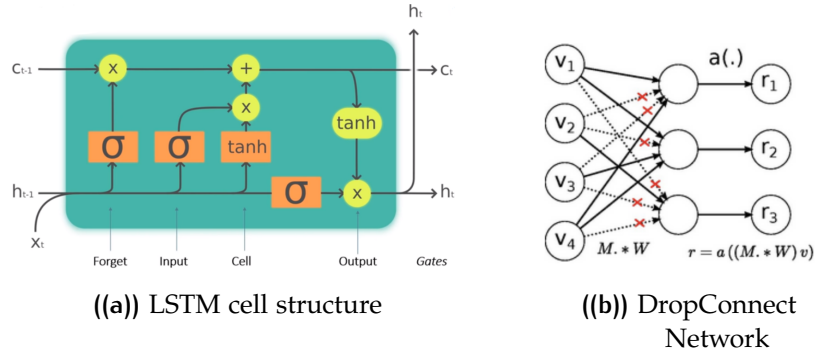


Figure 9: Models architecture

I will use the accuracy metric to access the relevance of the models. Accuracy is defined as follows :

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

where:

TP = True positive;

FP = False positive;

TN = True negative;

FN = False negative

Next step is to fine-tune this model and do transfer learning to create a new language model that's tailored specifically for predicting the next word of WSB comments. The learner has an accuracy of 32% which means the model is guessing the next word of the WSB comment correctly about a third of the time. Given the quality of the REDDIT data this is a decent result. For instance, if we feed the language model with "Chatter on the trading room floor is that TSLA will" and ask it to complete the sentence with ten words, it returns : "Chatter on the trading room floor is that TSLA will become a penny stock in the next few weeks ..". This is a proper and grammatically correct answer in my opinion. The part of the model that has the understanding of the sentence is called the encoder, and as such I saved it for the classification task.

Finally I trained the classifier passing the vocabulary from the language model and loading the encoding part. The model reached an accuracy of 67%. After turning the database upside down and playing with the hyperparameters, this is the best result I got. Given the relative weakness of this result and the difficulty I encountered in preventing my model from overfitting, we can see that the gamification

of trading alone struggles to explain the direction of the SP500 on the day following the WSB daily discussion threads. Nevertheless, the trading gamification needs to be factored into a broader model, as an alternative data source. The code source is on this [repo](#).

5 CONCLUSION

2021 might have been the year of the retail trader. Retail traders piled into meme stocks and options, launching asset prices like "rockets to the moon", and hopefully they had "diamonds hand" to wait for it to happen. While market commentators welcomed the resurgence of retail investors as a force in stock markets, online brokers did not unknowingly make active trading cheap, easy, and fun; they encouraged it. The grim reality is that brokers have a vested interest in fostering excessive trading by their clients. As a result, gamification of trading can be seen as a new driving force in the financial markets. However, using WSB's comments as a proxy for this phenomenon cannot be used alone in a predictive model on SPY. Yet regulation still has a say in the matter, so the question of how long gamification of trading will last needs to be asked.

REFERENCES

- [1] Andrade, F.R.H., Mizoguchi, R., Isotani, S. *The Bright and Dark Sides of Gamification*. https://link.springer.com/chapter/10.1007/978-3-319-39583-8_17. 2016.
- [2] Brian Burke. *Redefine Gamification to Understand Its Opportunities and Limitations*. https://blogs.gartner.com/brian_burke/2014/04/04/gartner-redefines-gamification/. 2014.
- [3] Chou, Y.K. *Actionable Gamification: Beyond Points, Badges, and Leaderboards*. Leanpub. 2016.
- [4] David Moberg Carl Moller. *Gamification of Stock Trading, Activating sleeping resources*. 2019.
- [5] Deterding, S., Dixon, D., Khaled, R., Nacke, L. *From Game Design Elements to Gamefulness: Defining Gamification*. https://www.researchgate.net/publication/230854710_From_Game_Design_Elements_to_Gamefulness_Defining_Gamification. 2011.
- [6] FINRA. *News Release*. URL: <https://www.finra.org/media-center/newsreleases/2019/finra-fines-robinhood-financial-llc-125-million-best-execution>.
- [7] Gregory W. Eaton et al. *Zero-Commission Individual Investors, High Frequency Traders, and Stock Market Quality*. 2021.
- [8] Hutto, C.J. Gilbert, E.E. *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. <http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf>. 2014.
- [9] James Fallows Tierney. *Gamification and securities regulation*. <https://ssrn.com/abstract=3916407>. 2021.
- [10] Jeremy Howard, Sebastian Ruder. *Universal Language Model Fine-tuning for Text Classification*. <https://arxiv.org/pdf/1801.06146.pdf>. 201.
- [11] Philipp Chapkovski, Mariana Khapko and Marius Zoican. *Does gamified trading stimulate risk taking?* 2021.
- [12] Matt Phillips and Taylor Lorenz. “‘Dumb Money’ Is on GameStop, and It’s Beating Wall Street at Its Own Game”. In: *The New York Times* (Jan. 27, 2021).
- [13] SEC. *Press Release*. URL: <https://www.sec.gov/news/press-release/2021-167>.