External Forces on Financial Markets: Evidence from the GameStop Short Squeeze and Flash Crash

Jeffrey Choy[a], Ben Wang[a], Abdullah AlShelahi[a], Romesh Saigal[a]

[a] *Department of Industrial & Operations Engineering, University of Michigan, Ann Arbor, Michigan, USA*

Abstract

In this research, we present a system of equations for measuring external forces in financial markets. As in physics, we claim that there is a measurable force caused by external market forces, which we call investor impatience and equate with gravitational force. We simulate this force and its accompanying energy conservation equation equation using a physics-based Eulerian fluid flow system. We test this idea by analyzing minute-by-minute data from meme stocks during a one-of-a-kind market event, the January 2021 the GameStop short squeeze. The resultant parameters show that external forces have an effect on stock prices. The investor impatience parameter is shown to improve out-of-sample forecasting of investor sentiment, estimated using comments from Reddit's WallStreetBets forum during the short squeeze. We expand our study to the 2010 flash crash, demonstrating that the system accounts for exogenous influences on market behavior.

Keywords: Short squeeze, Meme stock, GameStop, Flash crash, Econophysics, Market crash

1. Introduction

Markets are highly complex and interdependent, influenced by a multitude of factors. Exogenous events and unique market phenomena can dramatically destabilize prices. Such abnormalities are difficult to predict and, as is often the case, have unprecedented consequences. In addition, it is often challenging to decompose market price action into endogeneous components (e.g. changes in fundamental stock price valuations) and exogeneous components (e.g. liquidity cascades).

The GameStop short squeeze of January 2021 is a clear example of an exogeneous event on market prices [33]. Retail investors on social news website Reddit instigated a rally of GameStop's stock price [21]. The stock saw a 1,500% increase in price over a two-week period ending January 27 [21]. A subsequent -44% crash of the stock price occurred the next day. A variety of other so-called *meme stocks* simultaneously experienced similar price behavior, including those of AMC Entertainment Holdings, Inc., Bed Bath & Beyond Inc., and Eastman Kodak Company [31].

Another example is the flash crash of May 6, 2010, which resulted in steep drawdowns and recoveries of major stock indices within 36 minutes [19]. The Dow Jones Industrial Average sustained its second largest intraday point decline, dropping approximately 9% [19].

The idiosyncracies of each abnormality make risk management and portfolio allocation nontrivial. The flash crash was a market-wide phenomenon, whereas the GameStop short squeeze was isolated to a few stocks. While there are many models in the literature connecting market abnormalities, such as the COVID-19 pandemic, to markets, most do so on the mesoscopic level at highest. That is, the behavior of a relevant, yet

Email address: jeffchoy@umich.edu (Jeffrey Choy)

often small, subset of traded assets are examined under econophysical or econometric methods. In contrast, macroscopic models, which consider all traded assets in a market simultaneously, capture interdependencies between assets invisible to smaller-scale models. There are conceptual benefits for risk management in taking such a perspective.

This research addresses the aforementioned challenges by aiming to quantify external influences on markets macroscopically. To do so, we extend the AlShelahi and Saigal macroscopic model of equity markets, a physics-inspired model that treats each stock as a particle within an Eulerian fluid-flow system of stochastic partial differential equations [1]. Calibration of this model during the flash crash has indicated the model describes market abnormalities. Our contribution is the decomposition of stock price acceleration into an endogenous and exogeneous component, the latter of which we term *investor impatience*. We compare these external forces to an invisible gravitational field, applying the conservation of energy principle to estimate these forces.

We validate the investor impatience parameter captures market abnormalities with two notably exogenous events: the GameStop short squeeze and the flash crash. For the former, we use minute-by-minute U.S. equity data and WallStreetBets comment sentiment estimates to demonstrate the parameter improves out-of-sample forecasting of comment sentiment. In the latter case, we demonstrate the primary feature of the flash crash, the unified drawdown across the entire market, is also exhibited in the investor impatience parameter. We therefore demonstrate the macroscopic model is not only capable of detecting market-wide events, but also add context to events on smaller scales. This enables potentially broad applications of the model to portfolio management under unexpected events.

The rest of this paper is structured as follows. Section 2 lays out the related literature and the contribution of this model in researching market abnormalities. Section 3 describes our methods and data. Our results are in Section 4, followed by concluding remarks in Section 5.

2. Related Literature

Our research increments on a considerable body of literature examining the effects of world events on financial markets. Methods from multiple disciplines have been applied, including from econometrics, signal processing, and physics, among others. We review approaches used for this purpose. We also discuss the application of sentiment analysis methods in financial contexts.

The versatility of econometric models in identifying relationships between time series makes unsurprising their popularity for contextualizing financial market abnormalities. An extensive body of literature applies vector autoregression (VAR) and its developments to relate global events to financial time series. Umar et al. [34] quantifies the return and volatility connectedness between COVID-19 media coverage and segments of the non-fungible tokens market. They use the same TVP-VAR approach as Antonakakis et al. [3], which employs the Diebold and Yılmaz [10] spillover index approach to identify volatility transmission between oil prices and the stock prices of oil and gas companies. Diebold and Yilmaz [9] employs a generalized VAR approach to quantify the volatility spillover across U.S. stock and bond, foreign exchange, and commodities markets through the global financial crisis. Shahrestani and Rafei [28] applies the Markov switching VAR model to measure the impact of oil price shocks on the Tehran Stock Exchange.

The literature also presents many applications of generalized autoregressive conditional heteroskedasticity (GARCH) models for detecting spillover and contextualizing market abnormalities. Dungey and Renault [11] proposes a GARCH common features approach which is used to identify contagion during major events in the Asian currency markets, global financial crisis, and European sovereign debt crisis. The dynamic conditional correlation MGARCH model presented in [12] has been used to detect contagion via estimates of time-varying conditional correlations. Among others, this method is applied to the global financial crisis in Syllignakis and Kouretas [32] and Kim et al. [18] to detect spillover during the global financial crisis from the U.S. to European and emerging Asian financial markets, respectively. Vasileiou [35] applies asymmetry GARCH models during the GameStop short squeeze to provide evidence of the presence of the anti-leverage effect.

Many authors have developed physics-inspired models of financial markets under the umbrella term *econophysics*. Such methods are frequently used in conjunction with econometric models. For a review of econophysics methods and applications, we refer the reader to the work of Chakraborti et al. [7]. The literature presents models for the detection and analysis of abnormal market events. Wavelets allow analysis of co-movements of financial data across different time scales. Ranta [27] applies wavelets to detect contagion between major markets across decades. Evidence of contagion is found during the 1987 financial crisis, the Gulf War, and the global financial crisis. Beccar-Varela et al. [6] applies a wavelet methodology designed for geophysical data to contrast the Lehman Brothers collapse and the flash crash using minute-by-minute stock data from four companies. The authors compare the former event to a natural earthquake and the latter to a human-made explosion and conclude that events of the former type are more predictable. Siddiqui et al. [29] identifies co-movement on short time scales between major stock indices during the onset of the COVID-19 pandemic. Xing et al. [40] posits that crashes originate from changes to the underlying structure of the financial system described by the nonlinear potential function. They use a GARCH model to improve forecasting of returns during market crashes. Wavelets have also been applied to cryptocurrency bubble analysis: Fruehwirt et al. [13] identifies a structural change in relationships between cryptocurrencies towards interdependence after the 2017 Bitcoin price peak. Kumar and Anandarao [20] uses wavelet coherence to identify volatility spillover in cryptocurrency markets. Beyond only market data, Umar et al. [33] uses Twitter data, the put-call ratio, and short-sale volume in a wavelet coherence approach to study the relationship between GameStop returns and sentiment during the short squeeze. Log-periodic power law (LPPL) models have been used to describe market abnormalities, inspired by statistical physics and motivated by distinct groups of rational and irrational traders. Geraskin and Fantazzini [14] provides a summary of the development and application of these approaches, beginning with the original model description from Sornette et al. [30]. Wosnitza and Denz [39] describes how LPPL structures follow the development of CDS spreads for forty banks during the 2000 financial crash. Applying a LPPL model to cryptocurrency, Wheatley et al. [38] diagnose bubbles and crashes in Bitcoin prices.

The phenomenon of social media, combined with advancements in natural language processing capabilities, has enabled data-driven analysis of the relationship between market action and the sentiment of market participants. This has produced a burgeoning literature on sentiment analysis applied to markets. Yang et al. [42] used posts from the Chinese stock message board to examine the effect of investor panic on equity market crashes. They used text-mining computing tools and a classification model to construct firm-level sentiment and panic indices which could predict abnormal trading and stock market crashes. Similarly, Xu et al. [41] constructed three sentiment indices from Chinese social media, newspapers, and internet news, which have proven capable of improving forecasting of stock market returns. As with forum posts, Google Search Volume Indices have been used as an indicator of investor intent. Hsieh et al. [15] used this data as a proxy for information demand of retail investors and to identify herding behavior of these investors. Lyócsa et al. [24] also used Google search volume to forecast global stock price variation during the COVID-19 pandemic.

The centrality of social media during the GameStop short squeeze has resulted in multiple studies using social media data to analyze the role of sentiment in this event. Wang and Luo [37] applies the VADER sentiment analysis package Hutto and Gilbert [17] and a BERT transformer model Devlin et al. [8] to WallStreetBets comments in a range of classification models to predict price movements of the GameStop during the short squeeze. Long et al. [23] also apply the VADER package to WallStreetBets comments to quantify the relationship between sentiment and GameStop returns. Mancini et al. [25] apply the VADER package to model the dynamics of emerging consensus within WallStreetBets during the short squeeze, making a comparison between GameStop's stock price and the transition to homogeneous opinions.

Despite a plethora of models for detecting and forecasting abnormalities in markets, the literature lacks models approaching the problem with a macroscopic perspective. While current models capture isolated phenomena due to unique events, broader implications may be left unaddressed. The scope of recommendations along the lines of portfolio management, policy, or risk management may therefore be limited. Our model addresses this research gap by constructing a market-wide sensor for abnormalities. We conduct sentiment analysis on WallStreetBets comments in a similar vein to the aforementioned studies to demonstrate our model captures the sentiment-induced abnormality of the GameStop short squeeze. The following section details the methods with which we construct said sensor and the data used for model fitting and validation.

3. Methodology

3.1. The AlShelahi and Saigal Macroscopic Model of Equity Markets

As mentioned, we expand on the AlShelahi and Saigal [1] macroscopic model of equity markets, a description of which is provided here. The model uses an Eulerian fluid-flow description of markets in which each stock's position is represented as its price. Taking $x \in \mathbb{R}^+$ to be a particular price, we construct $\rho(x,t)$ to be the density of stocks at price x and time t. Letting $N(x,t)$ denote the number of stocks in price section $[x_1, x_2]$ (with $x_1 \leq x \leq x_2$) at time t, we obtain the following definition for density:

$$N(x,t) = \int_{x_1}^{x_2} \rho(x,t)\, dx. \tag{3.1}$$

Density contextualizes the magnitude of shocks on markets: a shock to prices with greater density implies a higher impact. We may contrast density in this model with its physical interpretation. Whereas density classically refers to mass per unit volume, in this model density becomes the stocks per unit price. The velocity $v_k(t)$ of stock k at time t, is defined as

$$v_k(t) = \lim_{\Delta t \to 0} \frac{p_k(t) - p_k(t - \Delta t)}{\Delta t}, \tag{3.2}$$

where $p_k(t)$ is the price of stock k at time t. This parameter is similar to drift in classical stochastic differential equation models of stock prices. Indeed, the model may be applied to logarithmically transformed prices. We note that our results are similar with and without such a transform, opting for the latter for simplicity.

The average velocity of stocks in price interval $[x_1, x_2]$ at time t can be expressed as

$$v(x,t) = \frac{1}{N(x,t)} \sum_{k\,:\,p_k(t) \in [x_1, x_2]} v_k(t), \tag{3.3}$$

as can the average squared velocity, denoted as

$$v^2(x,t) = \frac{1}{N(x,t)} \sum_{k\,:\,p_k(t)\in[x_1,x_2]} v_k^2(t). \tag{3.4}$$

As [1] describes, we may assume that stocks are neither created nor destroyed, and thus the number of stocks can only change from flowing across endpoints of the interval $[x_1, x_2]$. This is because stock splits, delistings, and IPOs occur rarely. We may therefore define the flux (rate of flow) of stocks at point (x,t) as $Q(x,t)$ with

$$Q(x,t) = \rho(x,t)v(x,t). \tag{3.5}$$

Flux intuitively measures the scale of an event on stock prices, increasing in both the number of stocks impacted and the rate of price changes.

As [1] details, the conservation of mass principle can be applied to obtain an expression for pressure as a function of flux and velocity:

$$P(x,t) = \alpha Q(x,t)v(x,t), \tag{3.6}$$

where α is a fixed parameter. This pressure is a result of the momentum with which stocks are moving in the price domain. They found the empirical value of α to be 0.3, which is carried forward to this analysis.

Acceleration may be defined for a stock in a similar fashion to velocity:

$$a_k(t) = \lim_{\Delta t\to 0} \frac{v_k(t) - v_k(t-\Delta t)}{\Delta t}, \tag{3.7}$$

resulting in the following expression of average acceleration:

$$a(x,t) = \frac{1}{N(x,t)} \sum_{k\,:\,p_k(t)\in[x_1,x_2]} a_k(t). \tag{3.8}$$

With the above model description, [1] derives and fits the conservation of mass and momentum equations. The following subsection extends this to derive the gravitational parameter and conservation of energy equations to estimate external forces on financial markets.

3.2. Formulating External Market Forces as Gravity

As detailed above, stock prices are influenced by a multitude of factors, endogeneous and exogeneous. We make the assumption that external forces on stock prices generated by these markets and other world events act similarly to gravity in physics. However, unlike our gravitational experience, we surmise the external force on markets can act in either direction to increase or decrease stock prices. We may conceive of two opposing masses, positive (bullish) and negative (bearish) investor impatience, each with their own gravitational field. The relative size of each determines the net investor impatience.

We define the net force on stocks as the sum of internal and external forces as per Newton's second law of motion:

$$F(x,t) = m(x,t)\,a(x,t) + m(x,t)\,g(x,t) \tag{3.9}$$

where F represents the net force on the stocks priced x at time t, each with mass m, internal acceleration a, and external acceleration g. The mass may be expressed as the product of the stock fluid density, ρ, and volume, V:

$$m(x,t) = \rho(x,t)V(x,t). \tag{3.10}$$

Similarly, the force term may be expressed as a product of pressure and area, A:

$$F(x,t) = P(x,t)A(x,t). \tag{3.11}$$

Combining the above equations, we obtain

$$\alpha Q(x,t)v(x,t)A(x,t) = \rho(x,t)V(x,t)a(x,t) + \rho(x,t)V(x,t)g(x,t). \tag{3.12}$$

As the price-domain equivalent of three-dimensional volume is simply a price interval, we take V (and A) to be unit values. Combining the formula above with the definition of flux, we obtain

$$\alpha v^2(x,t) = a(x,t) + g(x,t). \tag{3.13}$$

Defining gravity allows for an expression of the potential energy of the market, allowing us to analyze the degree to which energy is conserved. As with momentum, the degree to which conservation is obtained may provide utility in sensing abnormal market events [1]. We may define the potential energy of a particular stock as

$$E_k^{(P)}(t) = \rho(p_k(t),t)g(p_k(t),t)p_k(t) \tag{3.14}$$

where p_k is the price of stock k. The average potential energy in a particular price interval $[x_1, x_2]$ can therefore be defined as

$$E^{(P)}(x,t) = \frac{1}{N(x,t)} \sum_{k\,:\,p_k(t)\in[x_1,x_2]} E_k^{(P)}(t). \tag{3.15}$$

We likewise define the kinetic energy of the stock, assumed to have unit mass as

$$E_k^{(K)}(t) = \frac{1}{2}v_k^2(t). \tag{3.16}$$

Consequently, we can quantify the kinetic energy of the price interval as

$$E^{(K)}(x,t) = \frac{1}{N(x,t)} \sum_{k\,:\,p_k(t)\in[x_1,x_2]} E_k^{(K)}(t). \tag{3.17}$$

Total energy at this price interval may therefore be specified as

$$E(x,t) = E^{(K)}(x,t) + E^{(P)}(x,t). \tag{3.18}$$

Applying the Euler energy conservation equation in one dimension, and assuming that energy, density, and velocity are all differentiable, we thus have

$$\frac{\partial E(x,t)}{\partial t} + \frac{\partial}{\partial x}((E(x,t) + P(x,t))v(x,t)) = 0 \quad \forall x, t > 0. \tag{3.19}$$

The complete model consists of three stochastic partial differential equations representing the conservation of mass, momentum, and energy principles. The following subsection details how we fit the model to data from the short squeeze and flash crash.

3.3. Data

3.3.1. The Data

For analysis of the Gamestop short squeeze, we use minute-by-minute price data collected from *Yahoo! Finance* for 4494 stocks listed on U.S. exchanges. Each stock was included only if data were available for at least 75% of the period from January 21 through January 29 2021, resulting in an effective sample size of 1690 stocks. When data for a particular minute was unavailable, we imputed the price from the most recently available price. Due to the sparsity of stocks with prices over $100, we analyze only those below this threshold. We obtain similar data for the flash crash and are left with an effective sample size of 2853 stocks following the same data-cleaning procedure.

We use the Pushshift Reddit dataset to obtain all available WallStreetBets comments during market hours from January 21 through 29 [5]. We exclude comments that have been deleted or removed, contain URLs, or were authored by the AutoModerator, an automated moderation tool. We further pre-process the comments by substituting usernames with '@user', cleaning the text of escape sequences, and converting HTML entities to their respective characters. The dataset contains approximately 1.8 million comments made during market hours.

3.3.2. Discretization and Fitting

We apply the discretizing and fitting approach used in [1] for parameter estimation. A resolution of $1 and 1 minute is applied. Consistent with their approach, a linear regression model is fitted relating observed acceleration to squared velocity, estimating gravity for a price/time combination of these variables

$$a(x,t) = -g(x,t) + \beta\alpha v^2(x,t) + \sigma(x,t)\epsilon, \tag{3.20}$$

where $\epsilon \sim \mathcal{N}(0,1)$ is an error term. To estimate $g(x,t)$, the regression uses calculated values for a and v^2 in the 3×3 grid of prices $\{(x_i, t_j) : i \in \{x-2, x-1, x\}, j \in \{t-2, t-1, t\}\}$.

We define a series $g_k(t)$ to be the external acceleration from the perspective of stock k:

$$g_k(t) = g(x,t), \quad p_k(t) \in S_x. \tag{3.21}$$

This highlights the applicability of the macroscopic model to phenomena on lower scales. Individual stocks' price changes can be contextualized within the broader macroscopic context.

3.3.3. Parameterizing the Energy Conservation Equation

Although the conservation equation (3.19) right-hand side is zero at all points, a forcing term is required due to discretization error and uncertainty which disturb this equation. We therefore propose a stochastic forcing term as follows, in similar fashion to the model in [1]:

$$\frac{\partial E(x,t)}{\partial t} + \frac{\partial}{\partial x}((E(x,t) + \alpha Q(x,t)v(x,t)))v(x,t)) = z(x,t) \tag{3.22}$$

$$z(x,t) = l(x,t) + \eta(x,t)E(x,t) + \sigma_3(x,t)\frac{dW_3(x,t)}{dxdt} \tag{3.23}$$

Here, $l(x,t)$ is a deterministic function representing the mean inflow or outflow of the right-hand side of (3.23). The $\eta(x,t)$ term is likewise a deterministic function for the rate of reversion to the mean of the right-hand side. The $W_3(x,t)$ term is a Brownian sheet, a Gaussian stochastic process with mean 0 and covariance $\mathbb{E}\left(W_3(x_1,t_1)W_3(x_2,t_2)\right) = \min(x_1,x_2)\cdot\min(t_1,t_2)$ [36]. The $\sigma_3(x,t)$ term represents the volatility of the energy conservation process.

3.4. Estimating Sentiment from Reddit Data

3.4.1. Introduction

We apply two techniques to estimate the sentiment of WallStreetBets comments. The first is VADER, a lexicon- and rule-based sentiment analysis tool designed specifically for social media text [17]. We also apply a pre-tuned version of the RoBERTa transformer-based machine learning model [17, 22, 16]. We use the Twitter-RoBERTa-base-sentiment model from CardiffNLP to estimate comment sentiment [16]. The RoBERTa-base model is pre-trained on the English Wikipedia and BookCorpus datasets [22]. The Twitter-RoBERTa-base model is then tuned on a set of approximately 58 million tweets [4]. Tasks this model is trained for sentiment analysis, irony detection, and hate speech detection, among other tasks. Although a RoBERTa model tuned on WallStreetBets comments would improve performance, there are none available at the time of our analysis. We consider the RoBERTa-base model tuned on Twitter data appropriate for this analysis as both Reddit and Twitter are social media websites and are thus likely to share commonalities. The similarity between the Twitter tuning data and our Reddit data is corroborated by our observation that, despite the option of long-form comments on Reddit, approximately 89% of comments in our dataset are within Twitter's original character limit of 140 characters. However, we acknowledge the limitations of this approach. Despite both being social media websites, Twitter and WallStreetBets have unique cultures. WallStreetBets comments are often ironic, esoteric, and loaded with forum-specific references and inside jokes. The model may therefore misclassify some comments from WallStreetBets. Further analysis may be required to both evaluate the model's accuracy on WallStreetBets comments and, if required, tune a model for use on WallStreetBets comments. The following subsections detail how comment sentiment is estimated and forecasting procedure.

3.4.2. Sentiment Estimation using VADER

The VADER sentiment package returns a normalized, weighted composite score for each comment's sentiment between 1 (most positive) and -1 (most negative). We denote the sentiment score of comment i as $\bar{s}_i$ and the time the comment was posted by $\hat{t}_i$. We construct a minute-by-minute estimate of the overall sentiment of the WallStreetBets forum by calculating the mean sentiment of all comments posted in each minute. We denote the average sentiment at minute t, given by $S_{\text{VADER}}(t)$, to be

$$S_{\text{VADER}}(t) = \frac{1}{|\{i : \hat{t}_i \in [t, t+1)\}|} \sum_{i:\hat{t}_i \in [t,t+1)} \bar{s}_i. \tag{3.24}$$

3.4.3. Sentiment Estimation using the Twitter-RoBERTa-base Sentiment Model

The Twitter-RoBERTa-base model for sentiment analysis outputs values corresponding to the labels negative, neutral, and positive. Denoting $\bar{n}_i$ and $\bar{p}_i$ to be the softmax transformations of the negative and positive output values for comment i, respectively, we estimate the average sentiment at minute t, denoted by $S_{\text{roBERTa}}(t)$, to be

$$S_{\text{roBERTa}}(t) = \frac{1}{|\{i : \hat{t}_i \in [t, t+1)\}|} \sum_{i:\hat{t}_i \in [t,t+1)} (\bar{p}_i - \bar{n}_i). \tag{3.25}$$

3.4.4. Forecasting Comment Sentiment

We use the Bayesian time series forecasting and inference package Orbit to predict the WallStreetBets comment sentiment estimations [26]. To construct a smaller set of regressors from the investor impatience field, we take the average investor impatience across price subsets. We choose the price interval of $2 to $30, as these price sections generally contain at least 10-15 stocks per dollar. We construct four average investor impatience series, denoted by $\bar{g}_i(t), i \in \{1, 2, 3, 4\}$ and defined as

$$\bar{g}_i(t) = \frac{1}{7} \sum_{j=1}^{7} g(2 + 7(i-1) + j, t), \quad i \in \{1, 2, 3, 4\}. \tag{3.26}$$

On each trading day from January 21st through 29th, we apply Orbit's BackTester method on the Damped Local Trend (DLT) model using an expanding window with a linear global trend, minimum training window length of 120 minutes, a forecast length of 30 minutes, and an increment length of 30 minutes. The number of samples for each model fit is 500. We use the symmetric mean absolute percentage error (SMAPE) as the error metric, since Orbit's models outperform in terms of this metric compared to other candidate time series models [26]. For each sentiment metric and each trading day, two models are fitted: one DLT model for the sentiment metric, and one using the investor impatience series as regressors for the sentiment metric.

4. Results and Discussion

4.1. Introduction

We first qualitatively analyze the investor impatience parameter across the GameStop short squeeze and flash crash. A macroscopic perspective is applied to notable days during these events to demonstrate the utility of the model's scale and comment on the model's potential use cases. This is followed by contextualizing individual stocks' price action during the short squeeze to show applicability of the model to smaller scales. We also demonstrate the potential for the model to serve as a sensor of market abnormalities via the conservation of energy equation parameters during the flash crash. Lastly, we present results from sentiment forecasting to quantitatively confirm the model is capable of detecting market abnormalities.

4.2. Macroscopic Investor Impatience

Figure 1 below shows the investor impatience field for January 25th during the short squeeze. Also plotted are the prices of three meme stocks of interest: AMC Entertainment Holdings Inc., Bed Bath & Beyond, and Eastman Kodak Company. We note that the color scale limits have been adjusted for visual aid due to the presence of outliers. Meme stocks saw large price increases before sustaining sharp declines beginning around 10:45am. The investor impatience parameter captures the sharp mid-morning drawdown of the meme stocks. The vertical striations of color display unity of price action across multiple price sections, demonstrating the model's ability to capture market-wide phenomena.

The equivalent graph for during the flash crash is shown in Figure 2 below, with the crash onset highlighted by the green dashed line. Similar color patterns are observed. Notably, a presence of wave-like negative investor impatience striations are observed prior to the onset of the crash. This corroborates the findings of [1], which indicates the model parameters may be used as sensors of abnormal market activity.

Figure 3 below shows the investor impatience parameter during January 28th, as well as the aforementioned stocks' prices. Once again, we see the presence of negative investor impatience parameter values throughout the drawdown phases.

These figures indicate potential applications of this model to hedging and regulation. The unification of investor impatience across price sections appears to occur more strongly during drawdowns than rallies. This supports the empirical observation that correlations between equities are much greater for downside moves than upside moves [2]. One may therefore use historical estimates of investor impatience to inform portfolio construction. Furthermore, the ability of the model to act as a sensor for market abnormalities may be used by regulators within a broader crash detection framework to prophylactically restrict a crashing market.

We explore the meme stocks' positions within the investor impatience field in the following subsection.

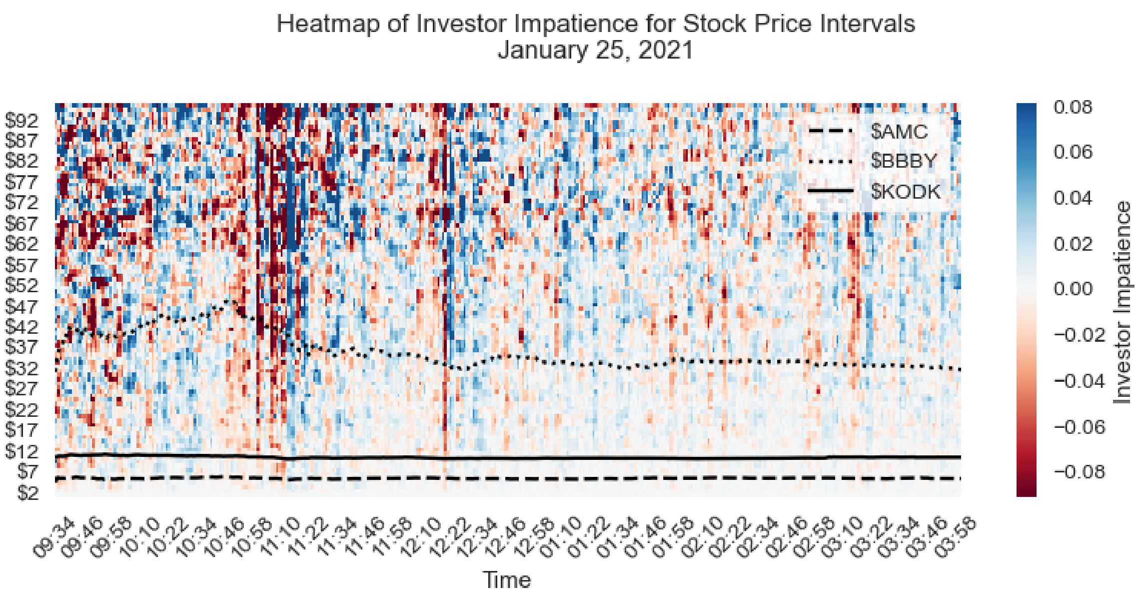

Figure 1: Investor Impatience Heatmap and Prices of $AMC, $BBBY, and $KODK for January 25, 2021

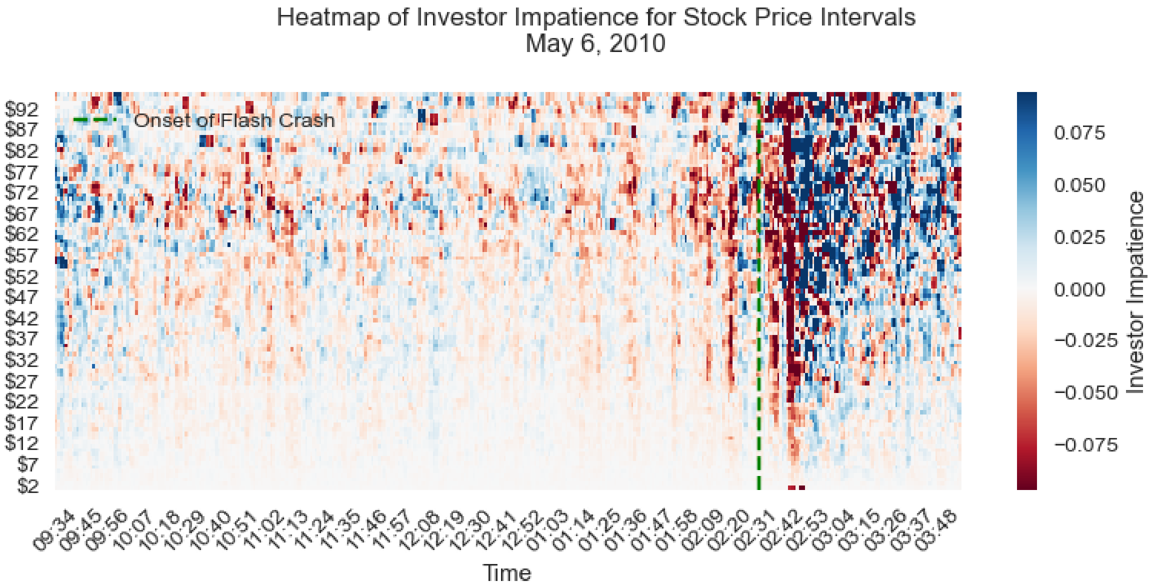

Figure 2: Investor Impatience for All Stocks during the Flash Crash

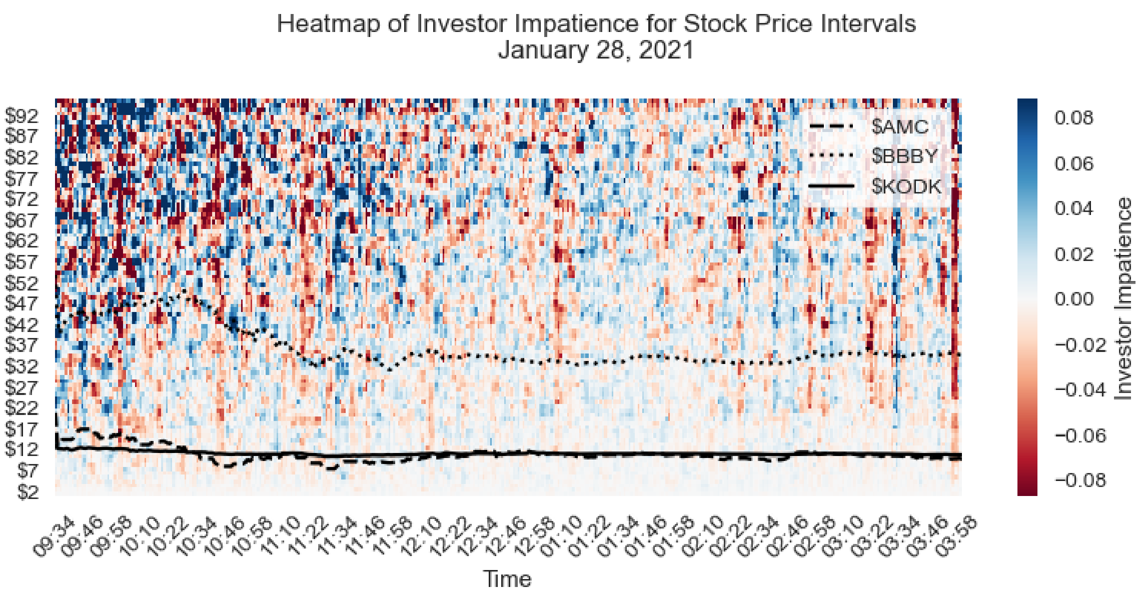

Figure 3: Investor Impatience Heatmap and Prices of $AMC, $BBBY, and $KODK for January 28, 2021

4.3. Microscopic Investor Impatience

We now examine the application of the model to individual stocks. Figures 4 through 6 below show the price, traded volume, and investor impatience parameter (i.e. $g_k(t)$) for Bed Bath & Beyond, AMC Entertainment Holdings Inc., and Eastman Kodak Company on January 28th. Also plotted is the 10-minute rolling average of investor impatience, for clarity. We note that GameStop's price was generally above $100 and is thus excluded. For all three stocks, considerable negative investor impatience is observed during the morning drawdown.

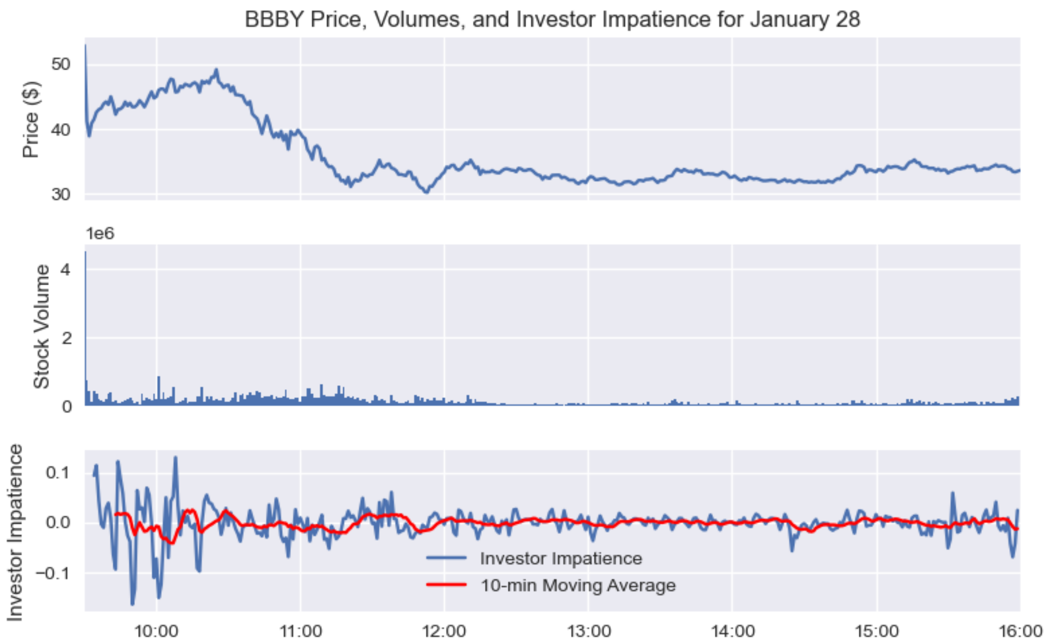

Figure 4: Bed Bath & Beyond ($BBBY) Price, Volume, and Investor Impatience for January 28, 2021

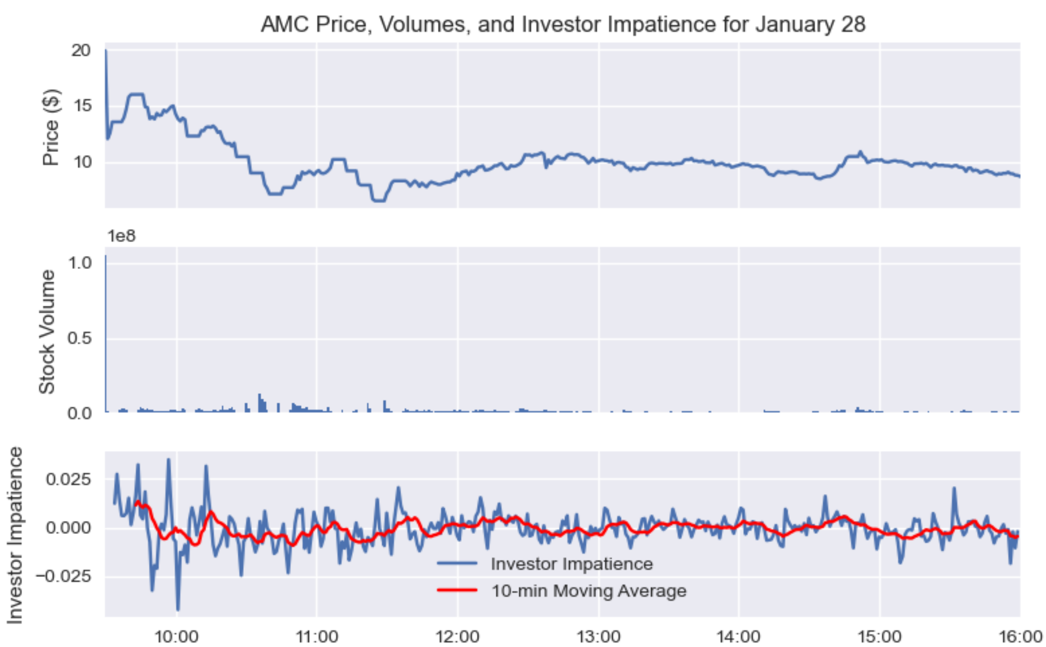

Figure 5: AMC Entertainment Holdings, Inc. ($AMC) Price, Volume, and Investor Impatience for January 28, 2021

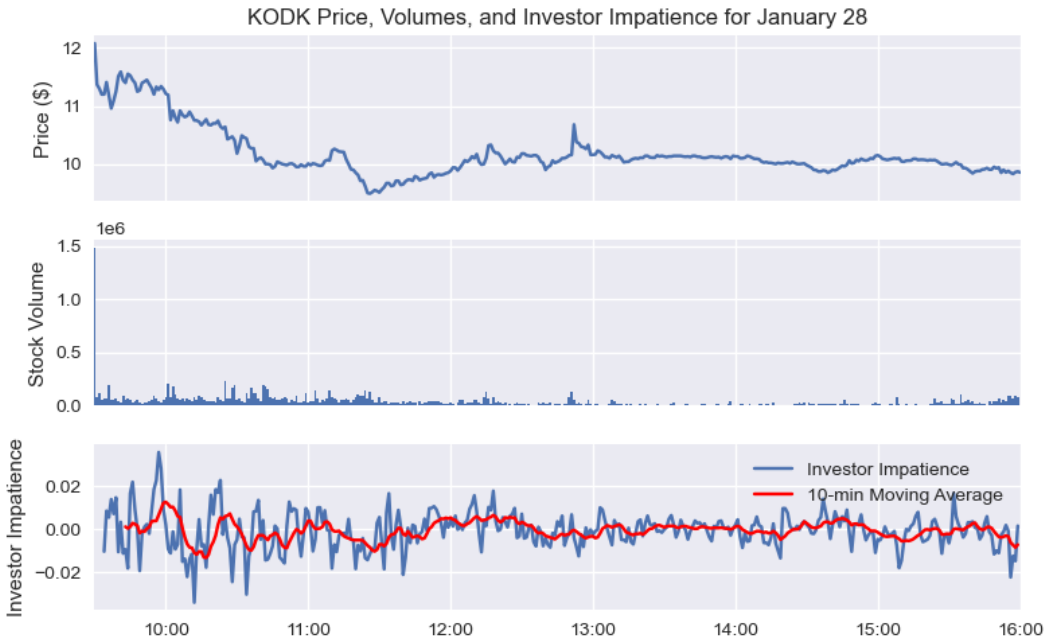

Figure 6: Eastman Kodak Company ($KODK) Price, Volume, and Investor Impatience for January 28, 2021

Figure 7 below displays the investor impatience and conservation of energy equation parameters during the flash crash. Again, the rolling 10-minute average of investor impatience is plotted. There is not only negative investor impatience prior to the crash, but also a noticeable increase in the energy conservation volatility term. These indicate this system may be used for detecting abnormalities in markets.

These figures also reveal how one may contextualize the price action of individual stocks within the investor impatience field to 'decompose' individual stock price action into components of market-wide external force and idiosyncratic action. Such an approach may be applicable to hedging. A stock with positive acceleration during a negative investor impatience enviroment (and vice versa) may have implications for portfolio diversification, particularly in market phenomena that impact a particular subset of traded assets.

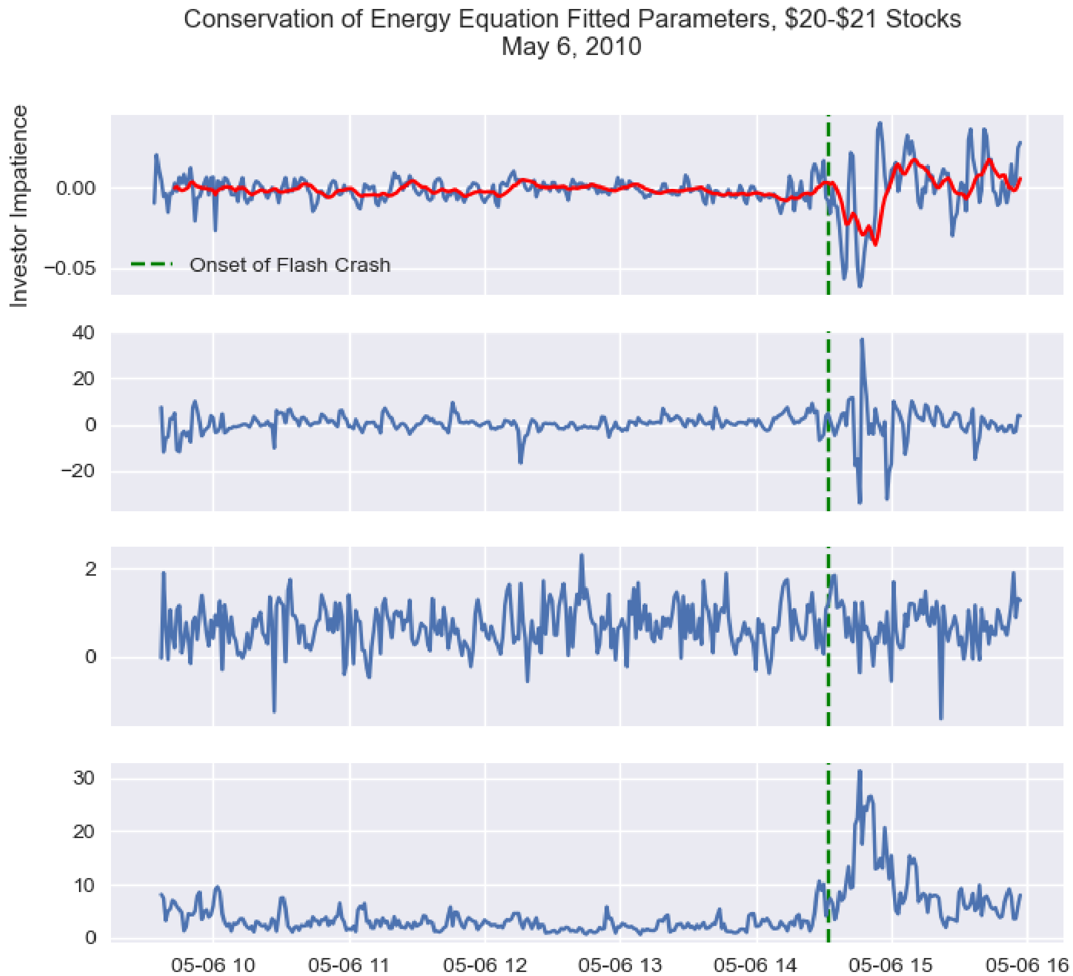

Figure 7: Investor Impatience and Conservation of Energy Parameters for $20 Stocks during the Flash Crash

4.4. Sentiment Forecasting

Table 1 below shows the out-of-sample SMAPE metric percentage improvement using the investor impatience regressors for each trading date using both sentiment models. Inclusion of the investor impatience regressors decreases the prediction error in five of seven days for both estimates of sentiment, although performance is superior for the Twitter-RoBERTa-base estimates. These results indicate the model parameters capture external phenomena that may have market impact, such as the role of investor sentiment in the short squeeze.

5. Conclusions and Future Work

We have extended AlShelahi and Saigal's physics-based macroscopic model of equity markets to include a gravitational term looking to capture external force on the market [1]. We fitted the investor impatience parameter and conservation of energy equation and analyzed them alongside a selection of stocks that exhibited abnormal behavior during the GameStop short squeeze. Our results indicate this model captures

	Sentiment Estimation Method	
Trading Day	**VADER**	**Twitter-RoBERTa-base**
January 21, 2021	-1.36%	-1.45%
January 22, 2021	-1.85%	2.38%
January 25, 2021	-5.49%	-4.18%
January 26, 2021	-5.52%	-0.37%
January 27, 2021	2.61%	-1.60%
January 28, 2021	8.99%	1.42%
January 29, 2021	-1.30%	-6.06%
Average	**-0.56%**	**-1.41%**

Table 1: SMAPE Prediction Error Change from Including Investor Impatience Regressors during the GameStop Short Squeeze

a degree of investor sentiment during this event and can be used as a sensor of abnormal market activity. The latter point is supported by our qualitative analysis of investor impatience and the energy conservation equation during the flash crash. Furthermore, the comparable results across contrasting market events indicates the ability of the model to respond to a potential multitude of external influences.

Multiple future avenues of research are presented by these findings. Naturally, future work may evaluate the investor impatience parameter and energy equation as predictive tools of market crashes. Similar analyses may be conducted on fixed income or cryptocurrency markets. Further analysis on the potential hedging capabilities of various portfolios from an investor impatience perspective are also warranted.

References

[1] A. AlShelahi and R. Saigal. Insights into the macroscopic behavior of equity markets: Theory and application. *Physica A: Stat. Mech. Appl.*, 505:778–793, 2018.

[2] A. Ang and J. Chen. Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63 (3):443–494, Mar. 2002. ISSN 0304-405X. doi: 10.1016/S0304-405X(02)00068-5.

[3] N. Antonakakis, J. Cunado, G. Filis, D. Gabauer, and F. Perez de Gracia. Oil volatility, oil and gas firms and portfolio diversification. *Energy Economics*, 70:499–515, Feb. 2018. ISSN 0140-9883. doi: 10.1016/j.eneco.2018.01.023.

[4] F. Barbieri, J. Camacho-Collados, L. Espinosa Anke, and L. Neves. TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1644–1650, Online, Nov. 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.148.

[5] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn. The Pushshift Reddit Dataset, 2020.

[6] M. P. Beccar-Varela, M. C. Mariani, O. K. Tweneboah, and I. Florescu. Analysis of the Lehman Brothers collapse and the Flash Crash event by applying wavelets methodologies. *Physica A: Statistical Mechanics and its Applications*, 474:162–171, May 2017. ISSN 0378-4371. doi: 10.1016/j.physa.2017.01.064.

[7] A. Chakraborti, I. M. Toke, M. Patriarca, and F. Abergel. Econophysics review: I. Empirical facts. *Quant. Financ.*, 11(7):991–1012, 2011. doi: 10.1080/14697688.2010.539248.

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018.

[9] F. X. Diebold and K. Yilmaz. Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1):57–66, Jan. 2012. ISSN 0169-2070. doi: 10.1016/j.ijforecast.2011.02.006.

[10] F. X. Diebold and K. Yılmaz. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134, Sept. 2014. ISSN 0304-4076. doi: 10.1016/j.jeconom.2014.04.012.

[11] M. Dungey and E. Renault. Identifying contagion. *Journal of Applied Econometrics*, 33(2):227–250, 2018. ISSN 1099-1255. doi: 10.1002/jae.2593.

[12] R. Engle. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002. ISSN 0735-0015.

[13] W. Fruehwirt, L. Hochfilzer, L. Weydemann, and S. Roberts. Cumulation, crash, coherency: A cryptocurrency bubble wavelet analysis. *Finance Research Letters*, 40:101668, May 2021. ISSN 1544-6123. doi: 10.1016/j.frl.2020.101668.

[14] P. Geraskin and D. Fantazzini. Everything you always wanted to know about log-periodic power laws for bubble modeling but were afraid to ask. *The European Journal of Finance*, 19(5):366–391, May 2013. ISSN 1351-847X. doi: 10.1080/1351847X.2011.601657.

[15] S.-F. Hsieh, C.-Y. Chan, and M.-C. Wang. Retail investor attention and herding behavior. *J. Empir. Financ.*, 59:109–132, 2020. ISSN 0927-5398. doi: 10.1016/j.jempfin.2020.09.005.

[16] Hugging Face. CardiffNLP/twitter-roberta-base-sentiment, 2022.

[17] C. Hutto and E. Gilbert. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proc. Int. AAAI Conf. on Web and Soc. Media*, 8(1):216–225, May 2014.

[18] B.-H. Kim, H. Kim, and B.-S. Lee. Spillover effects of the U.S. financial crisis on financial markets in emerging Asian countries. *International Review of Economics & Finance*, 39:192–210, Sept. 2015. ISSN 1059-0560. doi: 10.1016/j.iref.2015.04.005.

[19] A. Kirilenko, A. S. Kyle, M. Samadi, and T. Tuzun. The flash crash: High-frequency trading in an electronic market. *J. Financ.*, 72(3):967–998, 2017.

[20] A. S. Kumar and S. Anandarao. Volatility spillover in crypto-currency markets: Some evidences from GARCH and wavelet analysis. *Physica A: Statistical Mechanics and its Applications*, 524:448–458, June 2019. ISSN 0378-4371. doi: 10.1016/j.physa.2019.04.154.

[21] Y. Li. GameStop soars nearly 70%, trading briefly halted amid epic short squeeze, Jan. 2021.

[22] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. RoBERTa: A Robustly Optimized BERT Pretraining Approach, 2019.

[23] S. Long, B. M. Lucey, L. Yarovaya, and Y. Xie. ”I Just Like the Stock”: The Role of Reddit Sentiment in the GameStop Share Rally, Sept. 2022.

[24] Š. Lyócsa, E. Baumöhl, T. Výrost, and P. Molnár. Fear of the coronavirus and the stock markets. *Financ. Res. Lett.*, 36:101735, 2020. ISSN 1544-6123. doi: 10.1016/j.frl.2020.101735.

[25] A. Mancini, A. Desiderio, R. Di Clemente, and G. Cimini. Self-induced consensus of Reddit users to characterise the GameStop short squeeze. *Scientific Reports*, 12(1):13780, Aug. 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-17925-2.

[26] E. Ng, Z. Wang, H. Chen, S. Yang, and S. Smyl. Orbit: Probabilistic Forecast with Exponential Smoothing, 2020.

[27] M. Ranta. Contagion among major world markets: A wavelet approach. *International Journal of Managerial Finance*, 9(2):133–149, Jan. 2013. ISSN 1743-9132. doi: 10.1108/17439131311307556.

[28] P. Shahrestani and M. Rafei. The impact of oil price shocks on Tehran Stock Exchange returns: Application of the Markov switching vector autoregressive models. *Resources Policy*, 65:101579, Mar. 2020. ISSN 0301-4207. doi: 10.1016/j.resourpol.2020.101579.

[29] T. A. Siddiqui, H. Ahmed, and M. Naushad. Diffusion of COVID-19 impact across selected stock markets: A wavelet coherency analysis. *Investment Management and Financial Innovations*, pages 202–214, 2020.

[30] D. Sornette, A. Johansen, and J.-P. Bouchaud. Stock market crashes, Precursors and Replicas. *Journal de Physique I*, 6(1):167–175, Jan. 1996. ISSN 1155-4304, 1286-4862. doi: 10.1051/jp1:1996135.

[31] P. Stevens. AMC share price quadruples as retail traders raid hedge-fund short targets, Jan. 2021.

[32] M. N. Syllignakis and G. P. Kouretas. Dynamic correlation analysis of financial contagion: Evidence from the Central and Eastern European markets. *International Review of Economics & Finance*, 20(4): 717–732, Oct. 2011. ISSN 1059-0560. doi: 10.1016/j.iref.2011.01.006.

[33] Z. Umar, M. Gubareva, I. Yousaf, and S. Ali. A tale of company fundamentals vs sentiment driven pricing: The case of GameStop. *J. Behav. Exp. Financ.*, 30:100501, 2021. ISSN 2214-6350. doi: 10.1016/j.jbef.2021.100501.

[34] Z. Umar, A. Abrar, A. Zaremba, T. Teplova, and X. V. Vo. The Return and Volatility Connectedness of NFT Segments and Media Coverage: Fresh Evidence Based on News About the COVID-19 Pandemic. *Finance Research Letters*, 49:103031, Oct. 2022. ISSN 1544-6123. doi: 10.1016/j.frl.2022.103031.

[35] E. Vasileiou. Does the short squeeze lead to market abnormality and antileverage effect? Evidence from the Gamestop case. *Journal of Economic Studies*, ahead-of-print(ahead-of-print), Jan. 2021. ISSN 0144-3585. doi: 10.1108/JES-04-2021-0210.

[36] J. B. Walsh. An introduction to stochastic partial differential equations. In P. L. Hennequin, editor, *École d'Été de Probabilités de Saint Flour XIV - 1984*, pages 265–439, Berlin, Heidelberg, 1986. Springer Berlin Heidelberg. ISBN 978-3-540-39781-6.

[37] C. Wang and B. Luo. Predicting $ gme stock price movement using sentiment from reddit r/wallstreetbets. In *Proceedings of the Third Workshop on Financial Technology and Natural Language Processing*, pages 22–30, 2021.

[38] S. Wheatley, D. Sornette, T. Huber, M. Reppen, and R. N. Gantner. Are Bitcoin bubbles predictable? Combining a generalized Metcalfe's Law and the Log-Periodic Power Law Singularity model. *Royal Society Open Science*, 6(6):180538, 2019. doi: 10.1098/rsos.180538.

[39] J. H. Wosnitza and C. Denz. Liquidity crisis detection: An application of log-periodic power law structures to default prediction. *Physica A: Statistical Mechanics and its Applications*, 392(17):3666–3681, Sept. 2013. ISSN 0378-4371. doi: 10.1016/j.physa.2013.04.009.

[40] D.-Z. Xing, H.-F. Li, J.-C. Li, and C. Long. Forecasting price of financial market crash via a new nonlinear potential GARCH model. *Physica A: Statistical Mechanics and its Applications*, 566:125649, Mar. 2021. ISSN 0378-4371. doi: 10.1016/j.physa.2020.125649.

[41] Y. Xu, J. Wang, Z. Chen, and C. Liang. Sentiment indices and stock returns: Evidence from China. *Int. J. Financ. Econ.*, 2021. doi: 10.1002/ijfe.2463.

[42] X. Yang, Y. Zhu, and T. Y. Cheng. How the individual investors took on big data: The effect of panic from the internet stock message boards on stock price crash. *Pac.-Basin Financ. J.*, 59:101245, 2020.