Social Media Noise and Stock Manipulation

Douglas Cumming† and Vu Tran*

This version: April 15, 2025

Abstract

This paper models stock manipulations where investors interact via social network communications. We propose a novel noise index in social media platforms. The model predicts that high volumes of social media noise significant increase probability of success, profitability for manipulators as well as heighten trading volume of manipulated stocks. In addition, manipulation profitability increases with respect to the number of followers in social media posts mentioned the manipulated stock. Empirical investigations, based on over 3,800 U.S. small-cap stocks during January 2010 to 2018 December, confirm the theoretical predictions and hypotheses. Our paper demonstrates an urgent need for monitoring social media platforms in safeguarding financial market efficiency.

Keywords: Bias transmission, Market manipulation, Small Cap, Social media, StockTwits.

JEL Codes: G10, G11, G12, G40

† College of Business, Florida Atlantic University, Florida, US
* ICMA Centre, Henley Business School, University of Reading, Reading, UK

We are grateful for valuable suggestions and feedback from participants of ICMA research seminar. Remaining errors are of our own.

Social Media Noise and Stock Manipulation

This version: April 15, 2025

Abstract

This paper models stock manipulations where investors interact via social network communications. We propose a novel noise index in social media platforms. The model predicts that high volumes of social media noise significant increase probability of success, profitability for manipulators as well as heighten trading volume of manipulated stocks. In addition, manipulation profitability increases with respect to the number of followers in social media posts mentioned the manipulated stock. Empirical investigations, based on over 3,800 U.S. small-cap stocks during January 2010 to 2018 December, confirm the theoretical predictions and hypotheses. Our paper demonstrates an urgent need for monitoring social media platforms in safeguarding financial market efficiency.

Keywords: Bias transmission, Market manipulation, Small Cap, Social media, StockTwits.

JEL Codes: G10, G11, G12, G40

1. Introduction

Financial markets incorporate the wisdom of crowds while public opinions, in certain circumstances, can be biased via various mechanisms (Demarzo et al., 2003; Edmond, 2013). An example of such channels is transmissions of biases via social networks where (even professional) investors' decisions could be affected by their peers (e.g. Hong et al., 2004, 2005; Crawford et al., 2017; Han et al., 2021). Shiller (2014) suggests that social network communications appear to be an important determinant of stock market fluctuations. This paper incorporates social communications into the stock manipulation model, in Allen and Gale (1992) and Aggarwal and Wu (2006).

Our model shows that mass of misinformation and number of followers in social media platforms are critical to mislead small investors. Therefore, there should be strong associations between manipulators' profits and social media activities. In addition, we propose a novel noise index derived from social media posts. For the empirical validation of our model, we employ a comprehensive sample of all small cap stocks traded in NYSE and NASDAQ during the period from 2010 to 2018. The empirical investigations strongly support our model's predictions. We find that significant associations between manipulation profitability, trading volume and the noise index. In addition, both manipulation profitability and trading volume increases with respect to the number of followers in social media posts mentioned the manipulated stock.

Prior literature shows that humans are subject to a so-called "persuasion bias", that is, social influence of one agent on another's opinion formation depends not only on the accuracy of his/her signal(s) but also on how well-connected s/he is in social networks (e.g. Demarzo et al., 2003). In political literature, propaganda is explained by the magnification of repeated information in a similar logic. For example, Edmond (2013) models propaganda by repeated signals from multiple media outlets controlled by a (authoritarian) regime. Each citizen observes multiple noisy signals from the media outlets and eventually has biased beliefs. Allcott and Gentzkow (2017) show that misinformation diffuses effectively in modern social networks during the recent US Election. For example, pro-Trump fake news was shared 30 million times, far more than pro-Clinton ones. 14% of American adults consider social media as their most important source for election news. An obvious implication is that potential transmissions of biased beliefs are not restricted to important political circumstances.

Most models for speculative financial markets are analogous to Keynes's beauty contest, where participants try to predict the conventional consciousness of the true value not the true value itself. An asset's price, a reflection of the collective consciousness, could deviate from its fundamentals. Many prior studies suggest that investors are subjected to numbers of psychological biases and irrationalities (e.g. Hirshleifer and Teoh, 2003; Shiller, 2014).

We are not the first to examine biases transmissions via social networks. Hong et al., (2004) empirically show an association between investors' decisions in stock market participation and social interactions. Hong et al. (2005) reveals that fund managers are more likely to sell/buy financial assets that were sold/bought by their neighbours (in the same cities). They suggest that an epidemic pattern of information transmission through social networks exists even for institutional investors. Fund managers are shown to share investment ideas within their networks due to a number of reasons e.g. liquidity constraints and/or receiving feedbacks on their ideas (Crawford et al., 2017). Informed traders have incentives to spread imprecise signals, i.e. mixtures of noises and truth (e.g. Bommel, 2003), implying a long-term survival of noises in financial markets. Chen et al. (2014) show that articles and commentaries on a popular online forum, seekingalpha.com, predict future stock returns and earnings surprises. Han et al., (2021) show that self-enhancement and overconfident biases transmit through social network communications. Such transmissions alter investors' choices of active versus passive investment strategies and vice versa. Kogan et al., (2023) find that fraudulent news has a positive impact on returns and significant drops in trading and volatility afterwards. Allen et al., (2024) reveal a coordination role of social media platforms in cases of short sale squeezes. Hirshleifer et al. (2025) show that stronger social connections can facilitate faster incorporation of new information into market prices but also contribute to increased opinion polarisation and overactive and excessive trading activities.

Our work differs from prior literature in examining how biases and irrationalities transmit in financial markets. We focus on the cases of manipulated stocks. We extend Allen and Gale (1992) and Aggarwal and Wu (2006) model by incorporating social media communications and interactions between investors. Our model is also motivated by recent developments of social media and financial markets (e.g. Kogan et al., 2023; Allen et al., 2024).

The model predicts that high volumes of social media noise significant increase probability of success, profitability for manipulators. In addition, high social media noise is associated with heighten liquidity of manipulated stocks during the periods when manipulators exist the stock market. We propose a noise index derived from social media posts on related companies. For empirical validation of the model, we employ a sample of small capitalization companies listed in NASDAQ and NYSE during the period from January 2010 to December 2018. We harvest over 79 million social media posts mentioning any of the (over 3,800) sampled companies. The empirical investigations strongly support of the model propositions and hypotheses.

Taken together, the paper emphasises the impact of social communications, interactions, and information dissemination in asset pricing. In the context of stock manipulation monitoring and market efficiency safeguarding, the mass of misinformation can be vastly pumped up which leads to erratic behaviours in pricing and irrationality. Financial regulators should incorporate social media platforms in their monitoring tools. This study also raises a caveat against recent developments such as dropping of fact-checking mechanisms in social media platforms.

The rest of this paper is organised as follows: the next section proposes a model and testable hypotheses on social media and stock manipulation, section 3 describes data and the empirical framework, section 4 presents empirical results and section 5 summarises the conclusions.

2. Model and Hypothesis

We extend the discrete model of stock price manipulation in Allen and Gale (1992) and Aggarwal and Wu (2006). We stick with their assumptions and setups for the model with only 02 exceptions, as follows: (*i*) Aggarwal and Wu (2006) assume that all agents are risk-neutral while we assume that the small investors are risk-averse and informed investors and manipulators are risk-neutral; (*ii*) small investors' beliefs can be influenced by social media information.

It is noteworthy that the model is for small-cap shares. There are three types of investors in the economy, namely informed investors (superscripted ***I***), manipulators (superscripted ***M***), and uninformed information seekers (superscripted ***S***). It is reasonable to assume that informed investors and manipulators are large traders compared to the information seekers. Large traders usually hold diversified portfolios, hence, they can be assumed as risk-neutral agents, in the context of a (small-cap) company while information seekers are risk-averse.

Due to the economy of scale, informed investor ***I*** can acquire private information and knows whether there will be good or bad news. As a result, the stock value in the future will be high (V_H) or low (V_L). For simplicity, manipulator ***M***, on other hand, knows that there will be either no news or bad news. Otherwise, if s/he knows the future stock value will be high and choose to enter the market, s/he would be categorised as ***I***. Without loss of generality, each type of these traders can be replaced by an investor and all shares are of a company.

There are N small information seekers ($S_i, i \in [1, N]$). These relatively small investors are uninformed and seek out for information about whether the future stock price will be high or low. They can observe public information such as past prices and trading volume, and their beliefs can be influenced by rumours on social media platforms.

In addition, there is a market maker who simply stand ready to provide liquidity to the market. It is also reasonable to assume that the market maker is risk-neutral in the context of the small company's shares. The supply curve of the company's shares is defined as follows:

$$P(Q) = V_0 + bQ \tag{1}$$

where P is the market price of the share, Q is the quantity of the shares demanded by investors, and b is the slope of the supply curve.

We assume that initially all shares are held by the market maker. The initial price is V_0 where no one buy the share from the market maker. For completeness, we assume that the total shares outstanding are $\frac{V_H - V_0}{b}$ which infers that $P < V_H$. This implies that if investors wished to buy all the shares from the market maker, the price would be V_H.

The timing of the model is, as follows. At time 0, all shares are held by the market maker. At time 1, the informed investor and/or the manipulator enters the market with probabilities of ρ^I and ρ^M, respectively. By definition, the informed investor only enters the market if there will be good news, hence, the future stock value is V_H. It is equivalent to that the probability of V_H is ρ^I. The probability of V_L ,thus, is $1 - \rho^I$, when the informed investor does not enter the market. At time 2, the information seekers enters while the informed investor and the manipulator exists the market. All shares hold by the informed investor and/or the manipulator are transferred to the information seekers at this stage. At time 3, which represents the long-term equilibrium, the fundamental value of the share is revealed to be either V_H or V_L.

Given risk-neutrality of the market marker, the initial value of the share, at time 0, should be the expected value of the future fundamental value, at time 3.

$$V_0 = \rho^I V_H + (1 - \rho^I) V_L \tag{2}$$

2.1. There is no Manipulator in the economy

We start with a baseline model where there is no manipulator in the economy. At time 1, if the informed investor bought the company's shares, the information seekers could observe the shares being purchased. They know that the informed investor has good information about the firm's prospects, i.e. future price would be V_H. Although they are risk-averse, the prospects of V_H is certain under this baseline model. Each information seeker will demand a quantity $q_2^{S_i}$ of the shares at time 2 in order to optimise their utility function. This is equivalent to solving the follows:

$$\max_{q_2^{S_i}} (V_H - P_2^*)\, q_2^{S_i} \tag{3}$$

Where P_2^* is the price of the shares at the clearing condition.

The aggregate demand of all the information seekers at time 2 is

$$q_2^S = \sum_{i=1}^{N} q_2^{S_i} \tag{4}$$

The Equation 3 becomes:

$$\max_{q_2^{S_i}} V_H\, q_2^{S_i} - \left[V_0 + b \left(\sum_{i=1}^{N} q_2^{S_i} \right) \right] q_2^{S_i} \tag{5}$$

The market clearing condition at time 2 can be achieved by solving N first-order partial derivative conditions.

$$q_2^{S_i *} = \frac{V_H - V_0}{(N+1)b} \tag{6}$$

The aggregate demand from the information seekers is

$$q_2^{S_*} = \frac{N}{N+1}\frac{V_H - V_0}{b} \tag{7}$$

The clearing price at time 2 is

$$p_2^* = V_0 + b\left(\sum_{i=1}^{N} q_2^{S_i *}\right) = \mathrm{V}_0 + \frac{N}{N+1}(V_H - V_0) \tag{8}$$

Working regressively, the informed investor would demand a quantity q_1^* of the shares at time 1 by solving the following optimisation equation:

$$\max_{q_1^*} p_2^* q_1 - (V_0 + bq_1)q_1 \tag{9}$$

At time 1 quantity demanded by the informed investor is

$$q_1^* = \frac{N}{N+1}\frac{V_H - V_0}{2b} \tag{10}$$

and the price at time 1 is

$$p_1^* = V_0 + \frac{N}{N+1}\frac{V_H - V_0}{2} \tag{11}$$

The informed investor's profit is

$$\Pi^* = \frac{N^2}{(N+1)^2}\frac{(V_H - V_0)^2}{4b} \tag{12}$$

Each information seeker gains non-negative profit by simply following the informed investor.

$$p_2^* - p_1^* = \frac{N}{N+1}\frac{V_H - V_0}{2} > 0 \tag{13}$$

Under the assumption of no manipulator in the economy, this risk-free arbitrage opportunity explains the motivation for information seekers to participate in the trading activities. In the next subsection, we will relax this assumption and illustrate that risky arbitrage opportunity still offers strong explanation for information seekers to participate.

2.2. An Economy with Manipulators

In a more realistic scenario, there is a probability of that a manipulator imitates the informed investor, drives up prices of the shares. The information seekers continue to optimise their demands at time 2 based on their observations of the large trading activity at time 1. However, a key difference in this round is that their optimisations need to account for the risk of deceitful trading activity.

Aggarwal and Wu (2006) assume that the information seekers can observe the probabilities of truthful informed trader, manipulator, good and bad news. As a result, they can infer the probability that purchaser of the shares is the manipulator which are fixed at the posterior probability of the manipulator conditional on pooling strategies between the informed investor and the manipulator. We relax these assumptions. Information seekers are relatively small and uninformed investors. Therefore, they are unlikely to able to observe the unconditional probabilities of truthful informed trader (ρ^I) nor of manipulator (ρ^M) nor of good/bad news. Instead, the information seekers estimate the probability of that the manipulator entered the market conditional on that large trade took place in time 1, $\theta_i^S\big(Prob(M|I \cup M)\big)$.

Investors often interact with each others. We assume that an information seeker's belief is dependent on his/her network interactions. On the other hand, the informed trader and the manipulator do not update their beliefs. Recall, the informed trader (the manipulator) only enters the market at time 1 if there will be good (bad or no-) news. The informed trader knows the probability of the manipulator entering the market at time 1 and do not choose to participate in the network communications. Meanwhile the manipulator knows that the probability is 1, as s/he already entered the market at time 1, and purposely spread the opposite signal of 0. An information seeker i interacts with m agents in his/her network on the topics related to the company's prospects, recent large trade(s), and importantly update his/her belief. After one round of communications, this investor's belief post-communication is expressed by the following equation.

$$\theta_1^{Si} = \sum_{j=1}^{m} l_{ij}^1 \, \theta_0^{Sj} \tag{14}$$

Where

θ_0^{Si} and θ_1^{Si} are the estimation of the information seeker i before and after round 1 of communications, respectively.

l_{ij}^S is a psychological measure of how much information seeker i believes in seeker j. It is noteworthy that i might not interact with all agents in the network. l_{ij} can take zero if either s/he did not listen or does not trust signals from agent j. Therefore, l_{ij} subsumes the listening structure and weightings in agent i learning process in Demarzo et al. (2003).

$$l_{ij} \in [0,1)$$

$$\sum_{j=1}^{m} l_{ij} = 1$$

The updating rule in Equation (14) can be rewritten in vector notation, as follows:

$$\Theta_1 = L\Theta_0 \tag{15}$$

Where Θ_1 and Θ_0 is the vector of all information seekers' beliefs post- (prior-) communications.

Following prior papers, we assume that the information seekers are bounded-rational i.e., they are not able to distinguish new and repeat information.

L is the listening matrix which subsumes the listening structure and weightings of investors' beliefs on other agents' signals.

The dynamics of small investors' beliefs become:

$$\Theta_n = \left[\prod_{s=1}^{n-1} L_s\right] \Theta_0 \tag{16}$$

If the manipulator pumps misinformation to the public domain, s/he would utilise the periods leading to time 2 which are unlikely long-run periods. Therefore, we can set the listening matrix to be constant i.e. $L_s = L \;\; \forall s \in [1, n]$. In other words, over short-term periods when the manipulator utilise to spread misinformation e.g. weeks, months, information seekers are unlikely to update their listening vector.

Proposition 1

Under few reasonable assumptions, each information seeker's belief converges to a consensus over the probability of a manipulator entering the market at time 1 regardless of whether s/he initially believes that the manipulator has entered the market in time 1.

The assumptions for this proposition include:

(i) Information seekers are bounded rational i.e., they cannot separate brand new element and repeated signals in communications. This is quite reasonable since fully-rationality requires that an information seeker knows the entire listening matrix of the whole economy.

(ii) There is a set, A, of strongly connected information seekers in the economy (Demarzo et al., 2003). This means that information seekers have influence on each others.

$$\lim_{n\to\infty} \theta_n^{S_i} = w\Theta_0 = \theta \qquad \forall i \in [1, N] \tag{17}$$

Where the vector w is the solution for $wL = w$.

The information seeker i uses their estimation of the probability of a manipulator conditional on observing a purchase at time 1 to solve the following optimisation problem.

$$\max_{q_2^{S_i}} \left[E\left(W^{Si}\right) - \frac{1}{2}\gamma V\left(W^{Si}\right) \right] \tag{18}$$

Where

$E\left(W_2^{Si}\right)$ is investor i's expected payoffs

$V\left(W_2^{Si}\right)$ is investor i's expected variance of the payoffs

γ is coefficient for the investor's risk-averse.

This is equivalent to the following

$$\max_{q_2^{S_i}}\left[(1-\theta)(V_H-P_2^*)q_2^{S_i}+\theta(V_L-P_2^*)q_2^{S_i}-\frac{1}{2}\gamma\theta(1-\theta)(V_H-V_L)^2q_2^{S_i}\right] \tag{19}$$

Solving N first-order conditions, we have

$$q_2^{S_i^*}=\frac{(1-\theta)V_H+\theta V_L-V_0}{(N+1)b} \tag{20}$$

The aggregate demand is

$$Q_2^{S^*}=\frac{N}{N+1}\frac{(1-\theta)V_H+\theta V_L-V_0}{b} \tag{21}$$

The market clearing price at time 2 is

$$P_2^*=V_0+\frac{N}{N+1}[(1-\theta)V_H+\theta V_L-V_0] \tag{22}$$

The expected profit for the information seeker i is

$$\pi_i^{S*}=\frac{[(1-\theta)V_H+\theta V_L-V_0]^2}{(N+1)^2b} \tag{23}$$

Setting the aggregate quantity under the market clearing condition directly from Equation (21) to be strictly positive in order to avoid market breakdown, we have:

$$\theta<\frac{V_H-V_0}{V_H-V_L} \tag{24}$$

Information seekers enter the market only if their ex-ante estimations of risk, of the manipulator entering at time 1, are sufficient low.

Proposition 2

As long as (i) there is a subset B *of information seekers whose beliefs can be influenced by the manipulator; and (ii) there are enough amount of communications, the consensus on the probability of deceitful trading in time 1 (i.e.* θ*) is sufficiently small. This proposition comes directly from* ***Proposition 1****.*

If at least one information seeker considers the manipulator's signals, even with a small weight, the manipulator's misinformation, $\theta = 0$*, would be repeated and propagated many times via her/his 1-tier, 2nd-tier, 3rd-tier .. connections with other nodes in the social network. Eventually, information seekers always share a small estimation of the probability of a manipulator entering the market at time 1 regardless of their initial (un)certainty of the risk that manipulator has entered the market in time 1.*

The market clearing conditions at time 1 can be obtained by solving the following optimisation:

$$\max_{q_1}(p_2^* - p_1^*)\, q_1 \tag{25}$$

The aggregate quantity demanded at time 1 is

$$Q_1^* = \frac{N}{N+1}\frac{(1-\theta)V_H + \theta V_L - V_0}{2b} \tag{26}$$

and the market clearing at time 1 price is

$$P_1^* = V_0 + \frac{N}{N+1}\frac{(1-\theta)V_H + \theta V_L - V_0}{2} \tag{27}$$

The manipulator's expected profit is

$$\Pi^M = \frac{N^2}{(N+1)^2}\frac{[V_H - V_0 - \theta(V_H - V_L)]^2}{4b} \tag{28}$$

Proposition 3

The magnitude of the information seekers' consensus estimation (θ) *of the probability of the risk, that the manipulator enters the market conditional on large trade being observed at time 1, is a function of misinformation amount which they have consumed. This proposition comes directly from* **Proposition 2**.

$$\theta = \left(1 - \beta ln(\nu)\right)\frac{\rho^{M}}{\rho^{M} + \rho^{I}} \tag{29}$$

Where β *is a constant and* ν *is amount of misinformation that the information seekers consume prior to the point of making decisions, i.e. time 2.*

The first-order partial derivative of the information seekers' beliefs over the probability of that the manipulator entered the market is

$$\frac{\partial\theta}{\partial\nu} = -\frac{\beta}{\nu}\frac{\rho^{M}}{\rho^{M} + \rho^{I}} \tag{30}$$

Equation (30) shows θ *is a decreasing function with respect to the consumed volume of misinformation* (ν)*. The more misinformation/noise the information seekers consume, the less their consensus estimation of the likelihood of being misled by the manipulator.*

2.3. Hypotheses

Hypothesis 1: Equation (24) implies that manipulation is not risk-free. Manipulator could incur loses when buying the stock at time 1 with knowledge of future bad news and the information seekers did not enter the market at time 2, as their estimations of deceitful trading is high. It is critical for the manipulator to mislead the information seekers to a sufficiently small consensus estimation (θ) of the risk of being deceived conditional on observed large trade at time 1. Equation (30) shows that the estimation of the risk inversely depends on volume of noise/ misinformation in social network. Taken together, these equations infer that volume of noise/ misinformation determines successes of manipulations. This leads to Hypothesis 1 that successful manipulations are associated with high volume of noise in social media.

There are 02 important challenges for the empirical testing of this hypothesis, as follows. First, one has to define the criteria of (un)successful manipulations. We define a (un)successful manipulation where return from time 1 to time 2 is (non)positive. Secondly, what is the measure of noise in social media? We estimate the component of social media volume on a firm that is (i) uncorrelated to the firm's fundamentals such as size, revenues, profits, leverage etc., stock market condition, business cycles; and (ii) highly correlated to manipulation values. This estimation is employed to proxy for a firm-level social media noise index.

Hypothesis 2: Return of the manipulated shares from time 1 to time 2 increases with respect to the amount of communications / misinformation.

$$\Delta P_{t=2,t=1} = \frac{N}{N+1}\frac{(1-\theta)V_H + \theta V_L - V_0}{2} \quad (31)$$

Replacing θ from Equation (29) and rearrange the above Equation (31), we have

$$\Delta P_{t=2,t=1} = \frac{N}{2(N+1)}\left[(V_H - V_0) - (V_H - V_L)\big(1 - \beta ln(\nu)\big)\frac{\rho^M}{\rho^M + \rho^I}\right] \quad (32)$$

The first-order partial derivative of the return with respect to amount of misinformation is

$$\frac{\partial \Delta}{\partial \nu} = \frac{\beta(V_H - V_L)}{\nu}\frac{\rho^M}{\rho^M + \rho^I} \quad (33)$$

Equation (33) shows that the manipulator's profit is an increasing function with respect to the volume of noise/ misinformation. We utilise the volume of social media communication for the manipulated stock as the proxy for ν.

Hypothesis 3: Return of the manipulated shares from time 1 to time 2 increases with respect to the number of information seekers.

We can see that the first-order partial derivative of the return is positive.

$$\frac{\partial \Delta}{\partial N} = \frac{1}{(N+1)^2} > 0 \tag{34}$$

We use the numbers of followers of social media posts about the manipulated stock as the proxy for N.

Hypothesis 4: Combining Equations (21) and (30), we have the following.

$$\frac{\partial Q_2^*}{\partial v} = \frac{N}{N+1} \frac{V_H - V_L}{b} \frac{\beta(V_H - V_L)}{v} \frac{\rho^M}{\rho^M + \rho^I} > 0 \tag{35}$$

This equation shows that trading volume at time 2, i.e. when the manipulator exists the market, increases with respect to the volume of noise/ misinformation.

Hypothesis 5: From Equations (21) and (24), we have the following:

$$\frac{\partial Q_2^*}{\partial N} = \frac{1}{(N+1)^2} \frac{(1-\theta)V_H + \theta V_L - V_0}{b} > 0 \tag{36}$$

This equation shows that trading volume at time 2, i.e. when the manipulator exists the market, increases with respect to the mass of information seekers.

3. Data and Empirical Strategy

Our empirical investigations employ data of small cap stocks traded in NYSE and NASDAQ during the period from January 2010 to December 2018. We collect stock data from CRSP. The stock manipulation data is from SMARTS, Inc., and CMCRC (Capital Markets CRC). SMARTS and CMCRC compile data on potential stock manipulation incidents from more than fifty global stock exchanges. Their data is used by regulatory authorities in many countries. The stock manipulation data are industry measures of manipulation and were not created for the purpose of this study. We keep only companies that have market capitalization of less than 2 billion US dollars at the begin of the sample. Our sample includes 3,832 companies.

Social media data is from StockTwits platform. During the sampled period, we harvest over 79 million tweets that mentioned any of the sampled companies. We use the StockTwits API to collect all posts containing company ticker with the '$' symbol. For each tweet, we extract the tweet content, the time posed of the tweet, the name of the user, the number of likes, and the number of retweets. Following prior papers, we drop all non-English tweets. Additionally, special characters are deleted from tweet messages, such as link tokens (e.g., 'http', 'https', and 'www'), hashtag tokens (e.g., '#'), and user identifier tokens (e.g., '@'). All tweets containing only links or URLs or emojis are dropped. We end up with a dataset of about 25.89 million of tweets.

Table 1 presents the descriptive statistics of the social media activities, stock characteristics, and manipulation variables during the sample period. Our sample includes 296,878 stock-month observations for 3,832 small cap stocks from January 2010 to December 2018. The mean value of tweet numbers is 87.2 while corresponding numbers for bullish and bearish are 26.5 and 3.6, respectively. Social media posts tend to spread more bullish than bearish information on average. Median values for all tweets and bullish (bearish) tweets are 10 and 0, significantly lower than the corresponding mean values. These figures point out that many stocks have neither bullish nor bearish social media coverage on most days. In addition, numbers of followers / bullish followers / bearish followers are highly skewed. For example, the average number of followers is 10,640 but the median is just 517. On the market manipulation variables, the mean number of intraday manipulation alerts is 0.32 but the median is 0, suggesting manipulation is rare but occasionally intense and clustered.

A key task for our empirical tests is to identify time 1 and time 2. We use each suspicious flag to signify the begin of time 2. We use the one-year period before manipulation as time 0 and the one-year period after manipulation as time 3

We observe that manipulation alerts are clustered. For example, 75 percent of manipulation alerts during the sampled period followed other manipulation flag(s) within a calendar month. Therefore, we employ monthly data for the empirical tests. We also cluster adjacent months with any manipulation alert(s) on the same stock within a 3-month window into a single event. For example, there were manipulation flags in month 1 and month 3 and there was no manipulation during month 2. All the three months would be clustered into single event.

Our empirical investigations base entirely on the measurement of noise volume in social media. We estimate this via a 2-step approach, as follows:

Step 1: Regress social media volume on a firm by the firm's fundamentals, stock market condition, business cycles, and estimate the residuals.

$$V_{i,t} = \alpha + \beta F_{i,t} + \gamma M_{i,t} + \delta C_{i,t} + \epsilon_{i,t} \tag{37}$$

Where

$V_{i,t}$ is Social media volume on firm i at time t.[1]

$F_{i,t}$ is a vector of firm i fundamentals, including firm's size (i.e. log of total assets), market-to-book ratio, log-revenues, log-net income, leverage, current ratio.

$M_{i,t}$ is stock market return i.e. S&P 500 return.

$C_{i,t}$ is a vector of control variables including a set of year effects, month effects, industry effects.

Step 2: We regress the estimated residuals from Equation (37) $\hat{\epsilon}_{i,t}$ by logarithm of manipulation values.

$$\hat{\epsilon}_{i,t} = a + bManipulation_{i,t} + e_{i,t} \tag{38}$$

Where

[1] It is noteworthy that our social media volume is already purged of significant numbers of posts. The raw number of tweets on the sampled companies is over 79 million. We dropped over 53 million (67%) of them, mostly non-English, non-content e.g. only emojis / url links.

$\hat{\epsilon}_{i,t}$ is the estimated residuals from Equation (37).

$Manipulation_{i,t}$ is the suspicious trading value of firm i at time t.

We use the fitted values ($\hat{a} + \hat{b}Manipulation_{i,t}$) to proxy for volume of noise in social media. The measures of noise volume, then, is normalized to have a mean value of 100, followed Baker et al. (2021). We term this Noise Index.

For hypothesis 1, we conduct an event study of the noise index between successful and unsuccessful manipulations.

For hypothesis 2, we estimate:

$$\pi_{i,t} = \alpha + \beta Noise_{i,t} + \gamma C_{i,t-1} + u_{i,t} \tag{39}$$

Where

$\pi_{t,i}$ is manipulation profits on firm i at time t. This is the average of stock i return during the manipulation period.

$Noise_{i,t}$ is the Noise Index for firm i at time t.

$C_{i,t-1}$ is a vector of control variables including the (lagged) firm's characteristics, (lagged) market condition, and a set of a set of year effects, month effects, industry effects.

For hypothesis 3, we estimate:

$$\pi_{t,i} = \alpha + \beta Nfollowers_{i,t} + \gamma C_{i,t-1} + v_{i,t} \tag{40}$$

Where

$\pi_{i,t}$ is manipulation profits on firm i at time t. This is the average return of stock i during the manipulation period. We cluster adjacent months with manipulation alert(s) on stock i into a single event t. Therefore $\pi_{i,t}$ would be the average return of stock i over the event window.

$Nfollowers_{i,t}$ is the logarithm of followers of firm i in social media at time t.

$C_{i,t-1}$ is a vector of control variables including the (lagged) firm's characteristics, (lagged) market condition, and a set of a set of year effects, month effects, industry effects.

For hypothesis 4, we estimate:

$$Volum_{i,t} = \alpha + \beta Noise_{i,t} + \gamma C_{i,t-1} + \zeta_{i,t} \tag{41}$$

Where

$Volum_{i,t}$ is Trading volume for firm i at time t which total trading number of stock i traded during the month that manipulation alert(s) happened.

$Noise_{i,t}$ is the Noise Index for firm i at time t.

$C_{i,t-1}$ is a vector of control variables including the (lagged) firm's characteristics, market condition, and a set of a set of year effects, month effects, industry effects.

For hypothesis 5, we estimate:

$$Volum_{i,t} = \alpha + \beta Nfollowers_{i,t} + \gamma C_{i,t-1} + \zeta_{i,t} \tag{42}$$

Where

$Volum_{i,t}$ is Trading volume for firm i at time t which total trading number of stock i traded during the month that manipulation alert(s) happened.

$Nfollowers_{i,t}$ is the logarithm of followers of firm i in social media at time t.

$C_{i,t-1}$ is a vector of control variables including the (lagged) firm's characteristics, market condition, and a set of a set of year effects, month effects, industry effects.

4. Empirical Results

Estimation of Social media Noise Index

Table 2 reports regression results for Equations (37) and (38). Column (1) of Table 2 reports estimation results of Equation (37) where the dependent variable is logarithm social media volume and the independent variables includes a firm's fundamentals, market condition, and industry, year, month effects. It is not surprising that volume of tweets is strongly correlated with corporate fundamentals such as size, leverage, revenues, profitability as well as general stock market conditions, proxied by S&P 500 returns. However, our purpose for Equation (37) is not estimation of the relationship between the fundamentals and tweet volume, but to extract the component in tweet volume that is uncorrelated to the fundamentals, market condition, and business cycle i.e., the fitted residuals. This series is, then, used as the dependant variable in column (2) of Table 2. We estimate the relationship between abnormal tweet volume and suspicious trading values. Consistent with our theoretical model prediction, abnormal social media volume is strongly correlated with suspicious trading values at 0.1% significant level. The fitted values ($abn.\widehat{tweet.}volum$) from the estimation of Equation (38) measure the component in abnormal tweet volume that is strongly correlated to suspicious trading values. Following (Baker et al. (2021) method, we normalized this series ($abn.\widehat{tweet.}volum$) to have firm level social media noise index.

Noise Index and Manipulation Probability and Profitability

We test Hypothesis 1 using t-tests for the null hypothesis that there is no difference in abnormal social media volume, abnormal follower counts and noise index values between unsuccessful and successful manipulation cases. Recall, we define a (un)successful manipulation where return from time 1 to time 2 is (non)positive. Table 3 presents the results from these tests. All the measures of social media activities are strongly significant at 0.1% level. Abnormal tweet volume, abnormal follower count, noise index are all significantly lower in unsuccessful manipulation cases. For example, the abnormal tweet volume in unsuccessful cases is over 12% lower, on average. It is noteworthy that this measure already accounts for firm fundamentals such as size, revenues, profitability etc, stock market condition, business cycle. Similarly, the number of followers in unsuccessful cases is over 27% lower than the corresponding figure in unsuccessful ones, ceteris paribus. Unsuccessful manipulation is also associated with significant lower social media noise, about 5% lower on average.

Table 4 reports regression results for Equation (39) in testing of Hypothesis 2. The table presents regression results examining how the manipulators' profits are influenced by social media noise. In column (1), we control for industry, month and year fixed effects. In column (2), we add lagged firm characteristics namely size, market-to-book, revenues, etc. Model (3) include lagged return and lagged S&P 500 return. Model (4) controls for all the mentioned variables. We cluster observations within individual firm and year for estimations of covariance matrices in all columns. Across all models, the coefficient of the noise index is strongly significant and positive.[2] A strong and statistically significant relationship suggests that as the noise volume in social media increases, the manipulators' profits also increase. The finding is in line with recent work by Dhawan and Putnins (2023) who evidences a significant role of social media in 355 cases of pump-and-dump manipulation in cryptocurrency markets. Among control variables, the coefficient of S&P 500 returns is statistically significant and positive. General market conditions play the most important role for manipulators' profits. The finding is consistent with prior papers which highlights investors are prone to herding and over-confident, hence, more exposed to manipulations during hot markets (DeLong et al., 1990; Barber and Odean, 2001; Jiang and Sun, 2014; Shiller, 2014). The findings are in support of our theoretical model's proposition and its prediction on the relationship between manipulation profits and noise volume in social media platforms.

Mass of Informational Seekers

Table 5 reports regression results for Equation (40) in testing of Hypothesis 3 that the manipulators' profits increase with respect to the number of informational seekers. We use the logarithm of the total follower counts of firm *i* at time *t* as the key independent variable in this

[2] In addition, we include lagged stock return and have very similar results (see Appendices A1-A4 for robustness checks of the corresponding in Tables 4 - 7).

regression equation. Like Table 4, model (1) controls for industry, month and year fixed effects. Model (2) includes additional variables i.e. lagged firm characteristics namely size, market-to-book, revenues, etc. In addition, Model (3) includes lagged return and lagged S&P 500 return. Model (4) controls for all the mentioned variables. Across all specifications, the coefficient of the logarithm follower count is consistently and significantly positive. Consistent with the results in Table 4, the coefficient of S&P 500 returns is also significantly positive. Notably, magnitude of the impact from the follower count is larger than that of the general market condition. This result is economically meaningful. It is much easier for the number of followers count to be increased by a certain percentage than the S&P 500 to increase by the same percentage. For example, the median of the follower count is merely 517 while the mean value is 10,640 (about a 20,000% change from the median). The maximum of the follower count is over 2.5 million which is equivalent to a 490,992% change from the median. Overall, the findings are in line with our model's prediction on the relationship between manipulation profits and the mass of informational seekers in social media platforms. This is also agreement with findings in Dhawan and Putnins (2023) that a (large) number of followers in social media platforms has positive impact on manipulators' profits in the markets of cryptocurrencies.

Evidence from trading volume

Table 6 reports regression results for Equation (41) in testing of Hypothesis 4 about the association between trading volume during manipulation periods and social media noise. The dependant variable is the logarithm trading volume during months with manipulation alerts. The key independent variable of interest is the social media index. Like Tables 4 and 5, we control for industry, month and year fixed effects, firm characteristics, stock market condition i.e. lagged S&P 500 return. Covariance matrices are clustered within individual firm and year. Again, the coefficient of the noise index is consistently significant and positive, regardless of the specifications of the empirical models. The magnitude of the impact from social media noise is much larger than that of all other dependant variables, even stronger than those observed in Tables 4 and 5. A 1-percent increase in the social media noise index is associated with approximate 0.58% - 0.6% increases in trading volume during manipulation periods. This is a huge impact given the variation in the noise index in our sample. The min and median values of the noise index are 52.03 and 52.04, respectively. The mean and the max values are 100 and 433. These statistics suggests that as the noise volume stays relatively small during most of the sampled periods and surges vastly during certain periods. This is similar to the evidence in Allen et al. (2024) who show erratic social media activities on meme stocks during short-sale squeeze periods.

Table 7 reports regression results for Equation (42) in testing of Hypothesis 5 about the relationship between trading volume during manipulation periods and mass of informational seekers. The dependant variable is the logarithm trading volume during months with manipulation alerts. The key independent variable of interest is the logarithm number of followers plus one. Like previous tables, control variables include firm characteristics, stock market condition i.e. lagged S&P 500 return, industry, month and year fixed effects. Covariance matrices are clustered within individual firm and year. Across all columns, the coefficient of the number of followers is always strongly significant and positive. The magnitude of the impact from the number of followers is lightly larger than that of the noise index (in Table 6). A 1-percent increase in the number of followers is associated with approximately 0.67% - 0.71% increases in trading volume during manipulation periods. Again, this impact is ginormous given the number of followers easily varies multiple folds during the sampled periods. The impact could be translated to increases of multiple-hundreds percentage points in trading volume during manipulation periods.

Falsification investigation

Tables 8 and 9 report the estimations of falsification regressions. We feed 10,000 random manipulation events when there is no actual manipulation alert during the month. There is no significant coefficient for neither the social media noise index nor the number of followers. The results further validate the findings in previous tables. Our main findings that social media noise index and mass of informational seekers are associated with positive increases in manipulation profits and volume are robust against omitted variables.

Taken together, our empirical investigations strongly support of the model's propositions and predictions. Social media can serve an important role for manipulators in pump-and-dump schemes in many senses including probability of successful manipulation as well as profitability, and liquidity of manipulated stocks.

5. Conclusion

Social media communications, machine news reading, swift information dissemination, and algorithmic trading are among key distinctions of today's financial markets. Noises could be exacerbated via social media communications (e.g. Allen et al. 2024; Dhawan and Putnins, 2023). This paper revisits a very crucial topic in finance, i.e., stock manipulation, noise traders and asset pricing. We present a simple extension of the Aggarwal and Wu (2006) model by incorporating the impact of social media communications. We propose a novel noise index in social media platforms. Our model predicts strong associations between manipulation profitability, trading volume and the noise index. In addition, the model also predicts positive sensitivity of both manipulation profitability and trading volume to mass of informational seekers which is proxied by the number of followers in social media posts mentioned the manipulated stock.

Empirical investigations, based on a comprehensive dataset of over 56 thousand manipulation alerts on over 3,800 small cap stocks traded in NYSE and NASDAQ, strongly support our model's predictions. We find that 1-unit increase in the social media noise index is associated with about 0.07% (0.6%) increases in manipulators' profits (trading volume) which is highly meaningful as the noise index varies greatly during the sample period with a standard deviation of 100 units. Furthermore, there is strong linkages between the number of followers and both manipulators' profits and their trading volume. A 1-percent increase in the followers count is associated with around 0.12% (0.66%) increases in manipulators' profits (trading volume). Again, these are of immensely significance given that the number of followers can erratically surge with magnitude of multiple-thousand percentages during manipulation periods (e.g. Allen et al. 2024).

This paper raises important implications both academically and practically. First, social interactions affect investment decisions and a social approach for investment modelling. This is particularly true in fast moving transmissions of (mis)information in social media platforms. Both irrationality and the mass of irrational investors are dynamics that needs to be taken to account for in asset pricing. Second, there is a strong association between the social media noise and manipulator's success, profitability, trading volume. These suggest practical implications for policy makers, investment and corporate managers. It is of an urgent need for financial regulators to incorporate social media (mis)information in their monitoring tools. This study also raises a caveat against recent developments such as dropping of fact-checking mechanisms in social media platforms. The extensive volume of social media data should be made accessible for research purposes. This data, in turn, plays a crucial role in detecting patterns and guidance for policymakers and practitioners concerning all aspects of social interactions and activities.

References

Aggarwal, R. K. and G. Wu (2006). Stock Market Manipulations. *Journal of Business* 79, 1915–1953.

Allcott, H. And Gentzkow, M. (2017) Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31,211-236.

Allen, F. and D. Gale (1992). Stock-Price Manipulation. *Review of Financial Studies* 5, 503–529.

Allen, F., Haas, M.D., Nowak, E., and Tengulov, A. (2023) Squeezing Shorts through social media platforms. *Swiss Finance Institute* working paper.

Baker, S. R., N. Bloom, S. Davis, and T. Renault (2021). Twitter-derived measures of economic uncertainty. NBER working paper 21633.

Bommel, V.J. 2003. Rumors. *Journal of Finance*, 58, 1499-1519.

Dhawan, A. and Putnins, T.J. (2023) A New Wolf in Town? Pump-and-Dump Manipulation in Cryptocurrency Markets. *Review of Finance*, 935 – 975.

Chen, H., De, P., Hu, Y.J., and Hwang, B.H., 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, 27, 1367-1403.

Crawford, S.S., Gray, W.R., and Kern, A.E., 2017. Why do fund managers identify and share profitable ideas? *Journal of Financial and Quantitative Analysis*, 52, 1903 -1926.

DeLong, J.B., Shleifer, A., Summers, L., and Waldmann, R., 1990. Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98, 703–38.

DeMarzo, P. M., D. Vayanos, and J. Zwiebel (2003). Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics* 118, 909–968.

Edmond, C., 2013. Information manipulation, coordination, and regime change. *Review of Economic Studies*, 80, 1422-1458.

Han, B., Hirshleifer, D., and Walden, J., 2021. Social transmission bias and investor behaviour. *Journal of Financial and Quantitative Analysis*, 390 – 412.

Hong, H., Kubik, J.D., and Stein, J.C., 2004. Social interactions and stock-market participation. *Journal of Finance*, 59, 137-163.

Hong, H., Kubik, J.D., and Stein, J.C., 2005. The neighbour's portfolio: word-of-mouth effects in the holdings and trades of money managers. *Journal of Finance*, 60, 2801-2824.

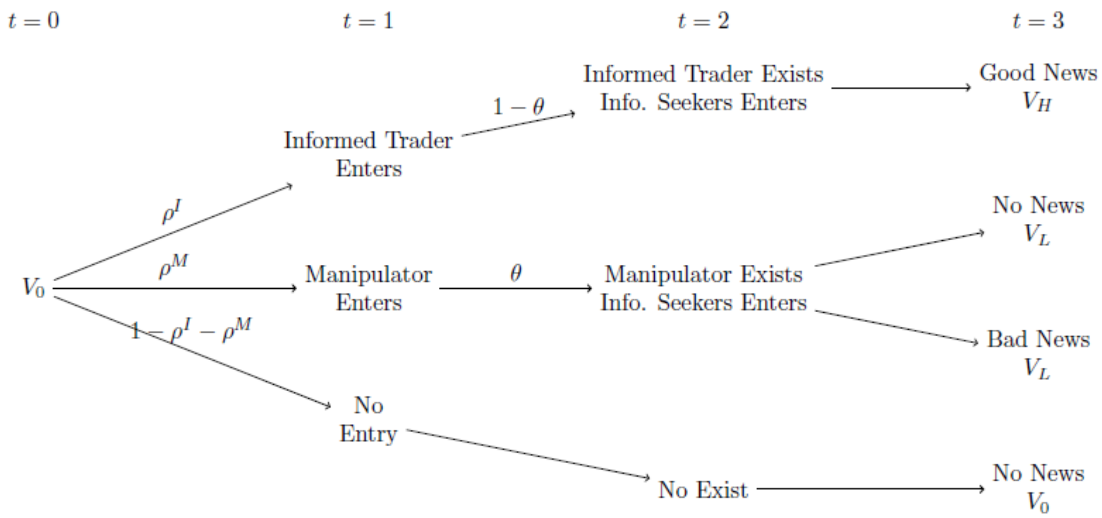
Hirshleifer, D., and Teoh, S., 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, 36, 337-386.

Hirshleifer, D. Peng, L., and Wang, Q. (2025) News diffusion in social networks and stock market reactions. *Review of Financial Studies* 38, 883–937.

Kogan, S., Moskowitz, T., and Niessner, M. (2023) Social media and financial news manipulation. *Review of Finance*, 1229-1268.

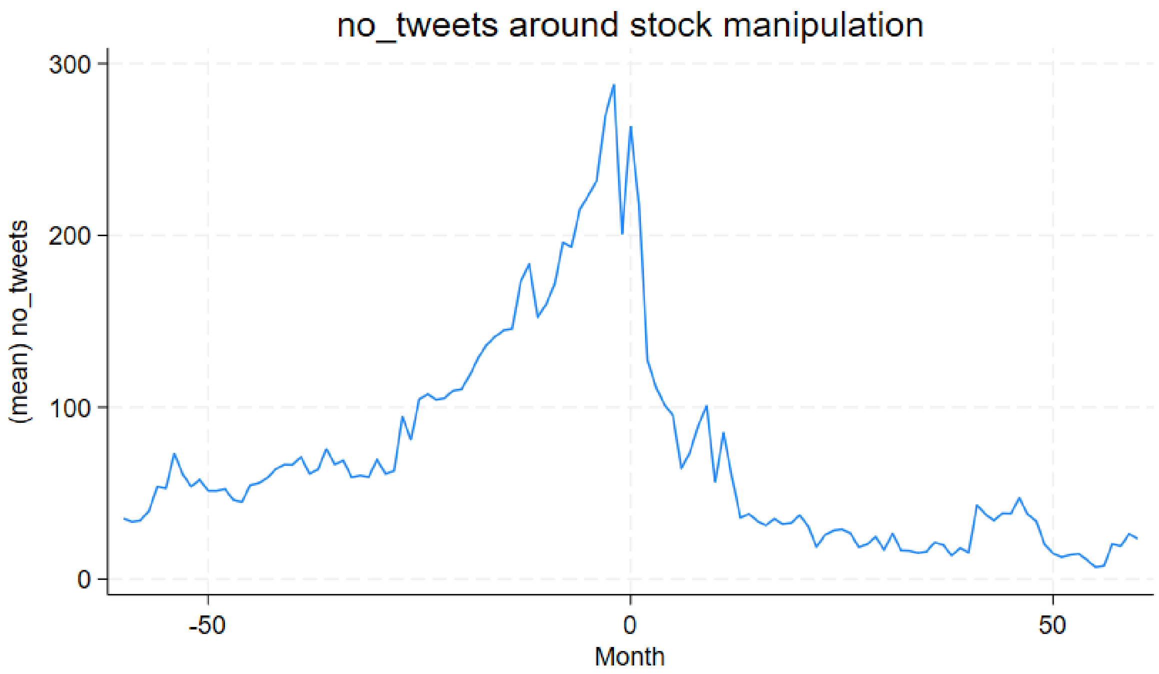
Shiller, R.J., 2014. Speculative asset prices. *American Economic Review*, 104, 1486-1517.

Figure 1. Timeline of Manipulation Events



Note: This figure depicts the timeline of events in our model for stock manipulation.

Figure 2. Social media volume around stock manipulation events



Note: This figure shows empirical observations of social media volume around manipulation events in our sample of (3,832) small cap stocks during 2010 - 2018.

Table 1: Summary statistics

	Mean	Std.Dev	Median	min	Max
Social media noise index	100	100	52.04	52.03	433.28
No. of Tweets	87.20	987.73	10.00	0.00	162,588
No. of Bullish Tweets	26.49	437.40	0.00	0.00	82,853
No. of Bearish Tweets	3.59	60.31	0.00	0.00	16,684
No. of Followers	10,640	37,666.64	517	0.00	2,538,433
No. of Bullish Followers	749.87	5,739.62	0.00	0.00	496,421
No. of Bearish Followers	142.61	1,439.32	0.00	0.00	279,102
Intraday Manipulation Alerts	0.32	0.91	0.00	0.00	26.00
Intraday Manipulation Value	345,297.22	2,591,052.54	0.00	0.00	4.01e+08
IntradayAlerts2Intervals Ratio	0.11	0.07	0.08	0.08	1.00
Intraday Manipulation2Turnover Ratio	0.21	0.15	0.17	0.00	0.99
EOD Manipulation Value	2,025.01	93,788.95	0.00	0.00	28,644,574
EOD Manipulation2Turnover Ratio (bpts)	3,332.38	2,532.99	2,688.48	0.72	10,000.00
Log Total Assets	6.15	1.69	6.29	-1.20	11.66
Log Market Cap.	5.83	1.52	6.00	-0.25	10.08
Market-to-Book	4.59	124.49	1.72	-4,164.71	14,450.66
Leverage ratio	0.28	0.40	0.13	0.00	6.27
Quick Ratio	3.23	3.43	2.25	0.01	83.60
Negative Book dummy	0.04	0.19	0.00	0.00	1.00
Negative Revenue dummy	0.00	0.02	0.00	0.00	1.00
Negative Net Income dummy	0.31	0.46	0.00	0.00	1.00
Return	0.01	0.15	0.00	-3.70	7.31
S&P 500 Return	0.01	0.03	0.01	-0.14	0.13
Observations	296,878				

This table reports the summary statistics for variables included in the empirical investigations. The sample includes monthly observations of 3,832 small cap stocks during the period from January 2010 to December 2018.

Table 2: Social media volume and stock manipulative values

	(1) Log.TweetsVolume	(2) Abn.TweetsVolume
Log Total Assets	0.1425***	
	(43.25)	
Market-to-Book	-0.0005	
	(-0.27)	
Leverage ratio	-0.0017	
	(-0.80)	
Quick Ratio	0.0107***	
	(5.12)	
Log Revenues	0.0030	
	(0.83)	
Log Net Income	-0.0277***	
	(-15.00)	
SP500 Return	0.0468***	
	(23.84)	
lmanipval		0.1811***
		(14.77)
Constant	***	***
	(12.89)	(-11.92)
Observations	192,522	192,522
Adjusted R-squared	0.369	0.033

This table reports the regression results in our 2-stage estimation of noise volume in social media. Stage (1), we estimate the component in social media volume which cannot be explained by company fundamental and stock market condition, and industry, year, month effects, i.e. the residuals from Equation (37). Stage (2), we estimate the fitted values from Equation (38) where the dependant variable is the outcome from Stage (1), and the independent variable is log- manipulative values. Standard errors are clustered at year, industry levels. T-statistics are reported in parentheses.

Table 3: Event study of social media activities

	beta	t-stat	p
Abn.TweetsVolume	-.1232487***	-11.19508	4.64e-29
Abn.Followers	-.274268***	-15.21118	3.82e-52
Noise Index	-5.430552***	-18.9876	3.84e-80
N	56,536		

This table reports the results for testing hypothesis 1. Specifically, we employ t-tests of the null hypothesis that there is no difference in social media activities between unsuccessful vs. successful manipulation cases. Abn.TweetsVolume is residuals from Equation (37) i.e., the component in (log) tweet volume that cannot be explained by company fundamental and stock market condition, and industry, year, month effects. Abn.Followers is residuals from the equivalent to Equation (37) where the dependant variable is replaced by Log- Follower Count. Noise Index is the measure of noise in social media i.e., uncorrelated to company fundamental and stock market condition, and industry, year, month effects, but strongly correlated to manipulative values.

Table 4: Sensitivity of the manipulator profits to social media noise

	(1) Manipul.Profits$_{i,t}$	(2) Manipul.Profits$_{i,t}$	(3) Manipul.Profits$_{i,t}$	(4) Manipul.Profits$_{i,t}$
Noise Index$_{i,t}$	0.0711***	0.0762***	0.0697***	0.0749***
	(8.68)	(5.38)	(8.39)	(5.27)
Log Total Assets$_{i,t-1}$		0.0127		0.0124
		(1.10)		(1.12)
Market-to-Book$_{i,t-1}$		0.0017		0.0013
		(0.38)		(0.26)
Leverage ratio$_{i,t-1}$		-0.0187**		-0.0188**
		(-2.51)		(-2.54)
Quick Ratio$_{i,t-1}$		0.0131		0.0127
		(1.16)		(1.15)
Log Revenues$_{i,t-1}$		-0.0316**		-0.0326**
		(-2.55)		(-2.59)
Log Net Income$_{i,t-1}$		0.0152		0.0147
		(1.45)		(1.37)
Negative Book$_{i,t-1}$		0.0021		0.0014
		(0.27)		(0.19)
Negative Revenue$_{i,t-1}$		0.0004		0.0004
		(0.44)		(0.38)
Negative Net Income$_{i,t-1}$		-0.0623***		-0.0625***
		(-5.77)		(-5.65)
S&P 500 Return$_{i,t-1}$			0.1105**	0.1095**
			(3.16)	(3.10)
Observations	54,039	41,511	54,039	41,511
Adjusted R-squared	0.028	0.036	0.039	0.046

This table reports the results for testing hypothesis 2. We estimate a regression of manipulator profits by social media noise. Control variables in Model (1) include industry, month and year fixed effects. Model (2) adds (lagged) firm characteristics namely size, market-to-book, revenues, etc. Model (3) include lagged S&P 500 return. Model (4) controls for all the mentioned variables. The t-statistics are reported in parentheses. We use covariance clustering within individual firm-year in all columns.

Table 5: Sensitivity of the manipulator profits to mass of information seekers

	(1) Manipul.Profits$_{i,t}$	(2) Manipul.Profits$_{i,t}$	(3) Manipul.Profits$_{i,t}$	(4) Manipul.Profits$_{i,t}$
Log-FollowersCount$_{i,t}$	0.1151*** (5.56)	0.1269*** (5.84)	0.1142*** (5.56)	0.1257*** (5.78)
Log Total Assets$_{i,t-1}$		0.0427*** (3.87)		0.0418*** (3.76)
Market-to-Book$_{i,t-1}$		0.0030 (0.44)		0.0026 (0.40)
Leverage ratio$_{i,t-1}$		-0.0240** (-2.92)		-0.0240** (-2.95)
Quick Ratio$_{i,t-1}$		0.0146 (1.31)		0.0142 (1.28)
Log Revenues$_{i,t-1}$		-0.0331** (-2.47)		-0.0341** (-2.50)
Log Net Income$_{i,t-1}$		0.0159 (1.52)		0.0153 (1.44)
Negative Book$_{i,t-1}$		0.0027 (0.31)		0.0021 (0.24)
Negative Revenue$_{i,t-1}$		0.0004 (0.90)		0.0004 (1.77)
Negative Net Income$_{i,t-1}$		-0.0690*** (-7.06)		-0.0691*** (-6.87)
S&P 500 Return$_{i,t-1}$			0.1111** (3.16)	0.1095** (3.08)
Observations	54,039	41,511	54,039	41,511
Adjusted R-squared	0.029	0.038	0.040	0.049

This table reports the results for testing hypothesis 3. We estimate a regression of manipulator profits by (log) number of followers. Control variables in Model (1) include industry, month and year fixed effects. Model (2) adds firm characteristics namely size, market-to-book, revenues, etc. Model (3) include S&P 500 return. Model (4) controls for all the mentioned variables. The t-statistics are reported in parentheses. We use covariance clustering within individual firm-year in all columns.

Table 6: Sensitivity of trading volume dumped by the manipulator to social media noise

	(1) Trading Volume$_{i,t}$	(2) Trading Volume$_{i,t}$	(3) Trading Volume$_{i,t}$	(4) Trading Volume$_{i,t}$
Noise Index$_{i,t}$	0.6081***	0.5766***	0.6076***	0.5769***
	(25.99)	(33.13)	(26.30)	(33.44)
Log Total Assets$_{i,t-1}$		0.0422		0.0423
		(1.52)		(1.54)
Market-to-Book$_{i,t-1}$		0.0119		0.0119
		(1.48)		(1.48)
Leverage ratio$_{i,t-1}$		0.0342*		0.0343*
		(2.13)		(2.14)
Quick Ratio$_{i,t-1}$		-0.0088		-0.0088
		(-0.60)		(-0.59)
Log Revenues$_{i,t-1}$		0.1370***		0.1372***
		(4.32)		(4.31)
Log Net Income$_{i,t-1}$		-0.0059		-0.0058
		(-0.75)		(-0.74)
Negative Book$_{i,t-1}$		0.0648***		0.0650***
		(4.01)		(4.02)
Negative Revenue$_{i,t-1}$		-0.0022		-0.0022
		(-0.98)		(-1.01)
Negative Net Income$_{i,t-1}$		0.2277***		0.2277***
		(21.41)		(21.46)
S&P 500 Return$_{i,t-1}$			-0.0178**	-0.0220***
			(-3.01)	(-4.19)
Observations	54,348	41,511	54,050	41,511
Adjusted R-squared	0.447	0.480	0.446	0.481

This table reports the results for testing hypothesis 4. We estimate a regression of logarithm trading volume during time 2 by (log) number of followers. Control variables in Model (1) include industry, month and year fixed effects. Model (2) adds firm characteristics namely size, market-to-book, revenues, etc. Model (3) include S&P 500 return. Model (4) controls for all the mentioned variables. The t-statistics are reported in parentheses. We use covariance clustering within individual firm-year in all columns.

Table 7: Trading volume dumped by the manipulator and mass of information seekers

	(1) Trading Volume$_{i,t}$	(2) Trading Volume$_{i,t}$	(3) Trading Volume$_{i,t}$	(4) Trading Volume$_{i,t}$
Log-FollowersCount$_{i,t}$	0.7194***	0.6650***	0.7127***	0.6652***
	(15.20)	(13.27)	(15.17)	(13.25)
Log Total Assets$_{i,t-1}$		0.3146***		0.3147***
		(8.36)		(8.37)
Market-to-Book$_{i,t-1}$		0.0261*		0.0262*
		(1.91)		(1.89)
Leverage ratio$_{i,t-1}$		-0.0159		-0.0159
		(-0.84)		(-0.84)
Quick Ratio$_{i,t-1}$		0.0131		0.0131
		(0.81)		(0.82)
Log Revenues$_{i,t-1}$		0.1240***		0.1241***
		(3.91)		(3.92)
Log Net Income$_{i,t-1}$		0.0001		0.0002
		(0.02)		(0.03)
Negative Book$_{i,t-1}$		0.0787***		0.0788***
		(3.92)		(3.93)
Negative Revenue$_{i,t-1}$		-0.0041		-0.0041
		(-1.06)		(-1.05)
Negative Net Income$_{i,t-1}$		0.1895***		0.1895***
		(17.35)		(17.34)
S&P 500 Return$_{i,t-1}$			-0.0118**	-0.0202***
			(-2.38)	(-4.32)
Observations	54,348	41,511	54,050	41,511
Adjusted R-squared	0.307	0.434	0.307	0.435

This table reports the results for testing hypothesis 4. We estimate a regression of logarithm trading volume during time 2 by (log) number of followers. Control variables in Model (1) include industry, month and year fixed effects. Model (2) adds firm characteristics namely size, market-to-book, revenues, etc. Model (3) include S&P 500 return. Model (4) controls for all the mentioned variables. The t-statistics are reported in parentheses. We use covariance clustering within individual firm-year in all columns.

Table 8: Placebo investigation: Manipulator profits with respect to social media noise

	(1) Manipul.Profits$_{i,t}$	(2) Manipul.Profits$_{i,t}$	(3) Manipul.Profits$_{i,t}$	(4) Manipul.Profits$_{i,t}$
Noise Index$_{i,t}$	0.0000	0.0000	0.0000	0.0000
	(.)	(.)	(.)	(.)
Log Total Assets$_{i,t-1}$		0.0694		0.0675
		(1.83)		(1.76)
Market-to-Book$_{i,t-1}$		0.0053**		0.0056**
		(2.46)		(2.69)
Leverage ratio$_{i,t-1}$		-0.0024		-0.0062
		(-0.18)		(-0.47)
Quick Ratio$_{i,t-1}$		0.0088		0.0103
		(0.43)		(0.49)
Log Revenues$_{i,t-1}$		-0.0475		-0.0450
		(-1.55)		(-1.44)
Log Net Income$_{i,t-1}$		0.0175***		0.0174***
		(5.15)		(4.89)
Negative Book$_{i,t-1}$		-0.0135		-0.0117
		(-0.59)		(-0.55)
Negative Revenue$_{i,t-1}$		0.0077**		0.0056*
		(3.10)		(2.27)
Negative Net Income$_{i,t-1}$		-0.0709***		-0.0712***
		(-5.36)		(-5.49)
S&P 500 Return$_{i,t-1}$			0.1166***	0.1354***
			(4.12)	(4.51)
Observations	8,659	6,214	8,659	6,214
Adjusted R-squared	0.025	0.039	0.037	0.055

This table reports falsification testing hypothesis 2. We seed placebo random manipulation events and estimate a regression of manipulator profits. Control variables in Model (1) include industry, month and year fixed effects. Model (2) adds firm characteristics namely size, market-to-book, revenues, etc. Model (3) include S&P 500 return. Model (4) controls for all the mentioned variables. The t-statistics are reported in parentheses. We use covariance clustering within individual firm-year in all columns.

Table 9: Placebo investigation: manipulator profits and mass of information seekers

	(1) Manipul.Profits$_{i,t}$	(2) Manipul.Profits$_{i,t}$	(3) Manipul.Profits$_{i,t}$	(4) Manipul.Profits$_{i,t}$
Log-FollowersCount$_{i,t}$	0.0681	0.0751	0.0619	0.0706
	(1.19)	(0.44)	(0.83)	(0.25)
Log Total Assets$_{i,t-1}$		0.0208		0.0200
		(0.62)		(0.60)
Market-to-Book$_{i,t-1}$		0.0070***		0.0069***
		(3.28)		(3.23)
Leverage ratio$_{i,t-1}$		0.0133		0.0137
		(0.65)		(0.67)
Quick Ratio$_{i,t-1}$		0.0238		0.0234
		(1.63)		(1.61)
Log Revenues$_{i,t-1}$		-0.0287		-0.0263
		(-1.10)		(-1.02)
Log Net Income$_{i,t-1}$		-0.0113**		-0.0085
		(-1.99)		(-1.44)
Negative Book$_{i,t-1}$		0.0095		0.0093
		(0.35)		(0.34)
Negative Revenue$_{i,t-1}$		-0.0249***		-0.0295***
		(-25.96)		(-27.20)
Negative Net Income$_{i,t-1}$		-0.0769***		-0.0759***
		(-5.47)		(-5.44)
S&P 500 Return$_{i,t-1}$			0.1155***	0.1334***
			(4.10)	(4.47)
Observations	86,39	6,139	8,639	6,139
Adjusted R-squared	0.016	0.020	0.034	0.036

This table reports falsification testing hypothesis 3. We seed placebo random manipulation events and estimate a regression of manipulator profits by (log) number of followers. Control variables in Model (1) include industry, month and year fixed effects. Model (2) adds firm characteristics namely size, market-to-book, revenues, etc. Model (3) include S&P 500 return. Model (4) controls for all the mentioned variables. The t-statistics are reported in parentheses. We use covariance clustering within individual firm-year in all columns.

Appendices

Table A1: Sensitivity of the manipulator profits to social media noise

	(1) Manipul.Profits$_{i,t}$	(2) Manipul.Profits$_{i,t}$	(3) Manipul.Profits$_{i,t}$	(4) Manipul.Profits$_{i,t}$
Noise Index$_{i,t}$	0.0711***	0.0762***	0.0390***	0.0464***
	(8.68)	(5.38)	(5.28)	(3.21)
Log Total Assets$_{i,t-1}$		0.0127		-0.0035
		(1.10)		(-0.22)
Market-to-Book$_{i,t-1}$		0.0017		-0.0014
		(0.38)		(-0.32)
Leverage ratio$_{i,t-1}$		-0.0187**		-0.0135**
		(-2.51)		(-2.37)
Quick Ratio$_{i,t-1}$		0.0131		0.0026
		(1.16)		(0.27)
Log Revenues$_{i,t-1}$		-0.0316**		-0.0166
		(-2.55)		(-1.58)
Log Net Income$_{i,t-1}$		0.0152		0.0134
		(1.45)		(1.42)
Negative Book$_{i,t-1}$		0.0021		-0.0008
		(0.27)		(-0.10)
Negative Revenue$_{i,t-1}$		0.0004		-0.0022*
		(0.44)		(-2.06)
Negative Net Income$_{i,t-1}$		-0.0623***		-0.0395***
		(-5.77)		(-4.62)
Return$_{i,t-1}$			0.3953***	0.3867***
			(21.33)	(20.71)
S&P 500 Return$_{i,t-1}$			-0.0241	-0.0195
			(-0.86)	(-0.69)
	*	*		
Constant	(-2.12)	(-2.03)	(-1.40)	(-1.82)

Observations	54039	41511	54039	41511
Adjusted R-squared	0.028	0.036	0.176	0.177

This table reports the results for testing hypothesis 2. We estimate a regression of manipulator profits by social media noise. Control variables in Model (1) include industry, month and year fixed effects. Model (2) adds (lagged) firm characteristics namely size, market-to-book, revenues, etc. Model (3) include lagged return and lagged S&P 500 return. Model (4) controls for all the mentioned variables. The t-statistics are reported in parentheses. We use covariance clustering within individual firm-year in all columns.

Table A2: Sensitivity of the manipulator profits to mass of information seekers

	(1) Manipul.Profits$_{i,t}$	(2) Manipul.Profits$_{i,t}$	(3) Manipul.Profits$_{i,t}$	(4) Manipul.Profits$_{i,t}$
Log-FollowersCount$_{i,t}$	0.1151***	0.1269***	0.0840***	0.0978***
	(5.56)	(5.84)	(4.52)	(4.80)
Log Total Assets$_{i,t-1}$		0.0427***		0.0118
		(3.87)		(0.91)
Market-to-Book$_{i,t-1}$		0.0030		-0.0010
		(0.44)		(-0.18)
Leverage ratio$_{i,t-1}$		-0.0240**		-0.0161**
		(-2.92)		(-2.63)
Quick Ratio$_{i,t-1}$		0.0146		0.0029
		(1.31)		(0.30)
Log Revenues$_{i,t-1}$		-0.0331**		-0.0175
		(-2.47)		(-1.63)
Log Net Income$_{i,t-1}$		0.0159		0.0137
		(1.52)		(1.46)
Negative Book$_{i,t-1}$		0.0027		-0.0010
		(0.31)		(-0.12)
Negative Revenue$_{i,t-1}$		0.0004		-0.0021***
		(0.90)		(-6.42)
Negative Net Income$_{i,t-1}$		-0.0690***		-0.0445***
		(-7.06)		(-5.82)
Return$_{i,t-1}$			0.3956***	0.3860***
			(21.63)	(21.34)
S&P 500 Return$_{i,t-1}$			-0.0239	-0.0193
			(-0.85)	(-0.68)
Constant	(0.32)	(-1.66)	(0.22)	(-1.60)
Observations	54039	41511	54039	41511
Adjusted R-squared	0.029	0.038	0.177	0.179

This table reports the results for testing hypothesis 3. We estimate a regression of manipulator profits by (log) number of followers. Control variables in Model (1) include industry, month and year fixed effects. Model (2) adds firm characteristics namely size, market-to-book, revenues, etc. Model (3) include lagged return and S&P 500 return. Model (4) controls for all the mentioned variables. The t-statistics are reported in parentheses. We use covariance clustering within individual firm-year in all columns.

Table A3: Sensitivity of volume dumped by the manipulator to social media noise

	(1) Trading Volume$_{i,t}$	(2) Trading Volume$_{i,t}$	(3) Trading Volume$_{i,t}$	(4) Trading Volume$_{i,t}$
Noise Index$_{i,t}$	0.6081***	0.5766***	0.6121***	0.5801***
	(25.99)	(33.13)	(26.79)	(33.58)
Log Total Assets$_{i,t-1}$		0.0422		0.0441
		(1.52)		(1.61)
Market-to-Book$_{i,t-1}$		0.0119		0.0122
		(1.48)		(1.51)
Leverage ratio$_{i,t-1}$		0.0342*		0.0337*
		(2.13)		(2.12)
Quick Ratio$_{i,t-1}$		-0.0088		-0.0076
		(-0.60)		(-0.52)
Log Revenues$_{i,t-1}$		0.1370***		0.1354***
		(4.32)		(4.30)
Log Net Income$_{i,t-1}$		-0.0059		-0.0057
		(-0.75)		(-0.73)
Negative Book$_{i,t-1}$		0.0648***		0.0652***
		(4.01)		(4.04)
Negative Revenue$_{i,t-1}$		-0.0022		-0.0019
		(-0.98)		(-0.85)
Negative Net Income$_{i,t-1}$		0.2277***		0.2251***
		(21.41)		(21.50)
Return$_{i,t-1}$			-0.0577***	-0.0443***
			(-8.11)	(-7.78)
S&P 500 Return$_{i,t-1}$			0.0018	-0.0072*
			(0.49)	(-2.04)
Constant	***	***	***	***
	(27.35)	(7.80)	(27.68)	(7.79)
Observations	54348	41511	54050	41511
Adjusted R-squared	0.447	0.480	0.449	0.482

This table reports the results for testing hypothesis 4. We estimate a regression of logarithm trading volume during time 2 by (log) number of followers. Control variables in Model (1) include industry, month and year fixed effects. Model (2) adds firm characteristics namely size, market-to-book, revenues, etc. Model (3) include lagged return and S&P 500 return. Model (4) controls for all the mentioned variables. The t-statistics are reported in parentheses. We use covariance clustering within individual firm-year in all columns.

Table A4: Sensitivity of trading volume to mass of information seekers

	(1) Trading Volume$_{i,t}$	(2) Trading Volume$_{i,t}$	(3) Trading Volume$_{i,t}$	(4) Trading Volume$_{i,t}$
Log-FollowersCount$_{i,t}$	0.7194***	0.6650***	0.7151***	0.6677***
	(15.20)	(13.27)	(15.23)	(13.34)
Log Total Assets$_{i,t-1}$		0.3146***		0.3174***
		(8.36)		(8.50)
Market-to-Book$_{i,t-1}$		0.0261*		0.0265*
		(1.91)		(1.93)
Leverage ratio$_{i,t-1}$		-0.0159		-0.0166
		(-0.84)		(-0.88)
Quick Ratio$_{i,t-1}$		0.0131		0.0141
		(0.81)		(0.88)
Log Revenues$_{i,t-1}$		0.1240***		0.1227***
		(3.91)		(3.89)
Log Net Income$_{i,t-1}$		0.0001		0.0004
		(0.02)		(0.05)
Negative Book$_{i,t-1}$		0.0787***		0.0791***
		(3.92)		(3.94)
Negative Revenue$_{i,t-1}$		-0.0041		-0.0039
		(-1.06)		(-0.99)
Negative Net Income$_{i,t-1}$		0.1895***		0.1874***
		(17.35)		(17.40)
Return$_{i,t-1}$			-0.0322***	-0.0336***
			(-4.82)	(-8.25)
S&P 500 Return$_{i,t-1}$			-0.0008	-0.0090
			(-0.13)	(-1.62)
Constant	***	***	***	***
	(52.30)	(13.23)	(53.80)	(13.29)
Observations	54348	41511	54050	41511
Adjusted R-squared	0.307	0.434	0.308	0.436

This table reports the results for testing hypothesis 4. We estimate a regression of logarithm trading volume during time 2 by (log) number of followers. Control variables in Model (1) include industry, month and year fixed effects. Model (2) adds firm characteristics namely size, market-to-book, revenues, etc. Model (3) include lagged return and S&P 500 return. Model (4) controls for all the mentioned variables. The t-statistics are reported in parentheses. We use covariance clustering within individual firm-year in all columns.