



**HAL**  
open science

## A Socio-Finance Model: The Case of Bitcoin

Yongqiang Meng, Dehua Shen, Xiong Xiong, Jørgen Vitting Andersen

► **To cite this version:**

Yongqiang Meng, Dehua Shen, Xiong Xiong, Jørgen Vitting Andersen. A Socio-Finance Model: The Case of Bitcoin. 2020. halshs-03048777

**HAL Id: halshs-03048777**

**<https://shs.hal.science/halshs-03048777>**

Submitted on 9 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CES

Centre d'Économie de la Sorbonne  
UMR 8174

**A Socio-Finance Model: The Case of Bitcoin**

Yongqiang MENG, Dehua SHEN  
Xiong XIONG, Jørgen Vitting ANDERSEN

**2020.31**



# A Socio-Finance Model: The Case of Bitcoin

Yongqiang Meng\*, Dehua Shen\*, Xiong Xiong\*, and Jørgen Vitting Andersen\*\*\*,\*\*\*\*

October 2020

\*College of Management and Economics, Tianjin University, Tianjin, 300072, China

\*\* CNRS, \*\*\*Centre d'Economie de la Sorbonne, Université Paris 1 Pantheon-Sorbonne,  
Maison des Sciences Economiques, 106-112 Boulevard de l'Hôpital, 75647 Paris  
Cedex 13, France.

## Abstract

This paper investigates the relations between multiple measures of investor sentiment and the returns, volatility, trading volume, and liquidity. Using both data outside and inside market, we find that the Bullishness from socio-finance model are significant related to future realized volatility and trading volume, similar to Tweet, which is thought to capture information of well-informed investors in Bitcoin market.

### 1. Introduction

In recent finance literature, Bitcoin, as the most popular cryptocurrency, has attracted increasing attention since it first introduced by Nakamoto in 2008. Bitcoin is considered as being lack of fundamental value, and the Bitcoin market is related with weak regulation, plenty of arbitrage opportunities, and high participation of retail investors, which give us an ideal place to investigate how investor sentiment influences asset pricing. As Baker and Wurgler (2007) state, the investigation of the effect of sentiment on assets is built on two broader and more irrefutable assumptions, which are sentiment and the situation of arbitrage. In order to capture the situation of investor sentiment and incorporate the influence of the trading characteristics on the situation of investor sentiment, we focus on multiple measures of investor sentiments, which is aim to find robust evidence of how investor sentiment influence Bitcoin pricing.

Having received such plenty of attentions, Bitcoin is therefore susceptible to attracting noise traders. As Qiu and Welch (2006) claim, under the conditions that assets held predominantly by noise traders and transaction costs are high enough to prevent systematic arbitrage, noise trader sentiment can persist and influence price in financial markets. The habitat of retail investors which are highly coincident with noise traders affects the relation between sentiment and return (Kumar and Lee, 2006; Baker and Wurgler, 2006; Li and Yeh, 2011). To capture the retail investor sentiment, scholars focus

on the behaviors of retail investors, including opening accounts number (Huang et al. 2009; Ni et al. 2015) and internet information (Joseph et al. 2011; Kim and Kim, 2014). Combining with the security and anonymity, Bitcoin is also traded as an alternative to the traditional assets.

The use of internet data to evaluate investor trading behavior of Bitcoin has become an important source of academic research. The findings based on internet data does not however generate consistent results. However, on balance the results seem to suggest that internet data regarding Bitcoin performance contain predictive information about returns. For instance, using Google trend data, Eom et al. (2019) claim that investor sentiment regarding Bitcoin has a significant information value for explaining changes in Bitcoin volatility for future periods. However, by decomposing Bitcoin price into efficient and noise components, Ibikunle et al. (2020) find that high levels of attention are linked with an increase in uninformed trading activity in the market for bitcoin, while informed trading activity is driven by arbitrage rather than attention. Using social media data, Shen et al. (2019) find that the number of tweets is a significant driver of next day trading volume and realized volatility. Choi (2020) employs high-frequency data and find that tweets predict liquidity in the next five to ten minutes. However, Guégan and Renault (2020) find that the significant relationship between investor sentiment, measured by StockTwits, and returns of Bitcoin is disappears for lower frequencies.

In this paper, we combine outside twitter data and inside trading data to measure the investor sentiment and investigate the relation between investor sentiment and Bitcoin characteristics. This paper uses daily number of tweets about “Bitcoin” as outside data to measure the investor sentiment on one hand, on the other hand, we employ socio-finance model from Anderson et al. (2014) which incorporates trading data as a proxy for investor sentiment. In particular, we investigate the suitability of the Tweets as a proxy for firm-level investor sentiment, since this data source may provide additional information regarding investors’ views on Bitcoin (Shen et al., 2019). Moreover, the socio-finance model serves a new view to measure the investor sentiment in Bitcoin. We focus on the Bitcoin market for several reasons. First, as the most successful virtual currency, Bitcoin was originally a monetary asset without underlying assets that its objective evaluation is difficult to determine, which makes it a speculative asset with very high volatility and attracting plenty of retail investors. This suggests that the sway of opinions will likely bring about a high level of consensus in retail investor views, regarding the valuation of Bitcoin. Second, unlike the traditional financial assets, Bitcoin has the benefits of security and anonymity for investors because banks, governments, or organizations do not issue it. Bitcoin market is characterized by continuous trading mechanism. These factors, make the Bitcoin market a market with lots of noise trades, which means that socio-finance model based on investors communications could capture investor sentiment from the trading data.

Our results can be summarized as follows. First, we find that our new investor sentiment proxy, Bullishness, predict future realized volatility and trading volume. Second, our results show that the number of tweets on Bitcoin can predict liquidity,

realized volatility and trading volume, but google trend index is caused by realized volatility and trading volume, which are consistent with previous literature (Urquhart, 2018; Shen et al., 2019). Third, as Shen et al. (2019) argue, the searching data from Google search engine is a good measure of attention from uninformed individuals who want to find out more information about Bitcoin, however, well-informed investors, who have knowledge of the cryptocurrency, will prefer tweet about the Bitcoin. Our new measure derived from inside market data is consistent with the twitter measure and proves that informed investor sentiment are associated with lower future realized volatility and trading volumes.

Our results contribute to prior results in the following ways. First, our results introduce a new measure of informed investors sentiment in Bitcoin. Second, we prove that the Bullishness is positive related the realized volatility and trading volume, which is also consistent with the relation between Tweet and Bitcoin dynamics.

The rest of this paper is structured as follows. Section 2 describes the data used in the study. Section 3 presents the empirical findings and section 4 summarizes and concludes the paper.

## 2. Literature review

### 2.1 Retail investors and investor sentiment in traditional financial assets

The behavioral and cognitive biases of investors, especially retail investors, are summarized to study the financial markets. Barber and Odean (2008) find that individual investors are more likely to be net buyers of attention-grabbing stocks indicated by daily abnormal trading volume, daily returns and daily news, than are institutional investors. Baber et al. (2009) also documents individual investors herd and small trade order imbalance reliably predicts returns in the opposite direction.

It has widely proven in different markets and different assets that investor sentiment has significant influence on assets pricing. Qiu and Welch (2004) compare investor sentiment measures based on consumer confidence surveys with measures extracted from the closed-end fund discount (CEFD) and find that sentiment plays a role in financial markets, but that the CEFD may be the wrong measure of sentiment. Ryu et al. (2016) find that investor sentiment generally plays a more important role in explaining stock market returns than investor trading behavior in Korean market. Gao and Yang (2017) find that high investor sentiment strengthens the positive relation between sentiment-returns in China future market. Ni et al. (2015) employ the panel quantile regression model and find the nonlinear (asymmetric and reversal) effect of investor sentiment on monthly stock returns in China stock market. Kumar and Lee (2006) find that systematic retail trading explains return movements for stocks with high retail concentration, especially if these stocks are also costly to arbitrage. Huang et al. (2009) build an investor sentiment index for China stock market using principal component employ a theoretical analysis about the relationship between investor sentiment and spot return.

### 2.2 Media information serves as investor sentiment proxy

The measurements of the investor sentiment of most literature are market wide (See e.g. Baker and Wurgler, 2006; Huang et al., 2009). In addition to the investor behaviors, scholars try to measure investor sentiment by media information.

One strain of the literature built the sentiment measurement on internet search. Da et al. (2011) measure investor attention using search frequency in Google (Search Volume Index (SVI)) and find an increase in SVI predicts higher stock prices in the next 2 weeks and an eventual price reversal within the year. Joseph et al. (2011) employ online ticker searches as a valid proxy investor sentiment and find its predictive ability of abnormal returns and trading volume in a weekly horizon for S&P 500 firms during 2005-2008. Dimpfl and Jank (2016) find bi-directional Granger causality of high searches measuring retail investor attention and high stock market volatility. Da et al. (2015) use daily internet search volume constructing an index to measure market-level sentiment.

Besides search information, traditional and social medias are another important source in measuring investor sentiment. Solomon (2012) finds that more media coverage increases announcement returns. Hillert et al. (2014) rely on 2.2 million articles from 45 US newspapers and find that firms particularly covered by the media exhibit, *ceteris paribus*, significantly stronger momentum. Zhang et al. (2016) find that the stock news on financial website have significantly relations with positive abnormal return and excessive trading volume on the event date. Social media sentiment is find to be associated with abnormal returns and trading volume (See e.g. Bollen et al., 2011; Sprenger and Welppe, 2014; Siganos et al., 2014).

Compared with internet search, traditional and social media, stock message boards attract mostly investors and document their sentiment timely. Wysocki (1998) use the sample on Yahoo! Message boards and find that cumulative posting volume is highest for firms with extreme past returns and accounting performance, high market capitalization, high price-earnings and market-to-book ratios, high volatility and trading volume, high analyst following and low institutional holdings. Other scholars also find that the investor sentiment proxy relying on internet postings have predictability on abnormal returns and trading volumes (See e.g. Tumarkin and Whitelaw, 2001; Antweiler and Frank, 2004; Leung and Ton, 2015). Furthermore, to increase the predictive accuracy, Zhang et al. (2012) contrast text classifiers to stock message board data to construct sentiment proxy. Li et al. (2018) find that Chinese internet stock message boards convey firm-specific information, which is represented by idiosyncratic volatility. By analyzing Chinese stock message boards, Huang et al. (2016) find that individual investors pay more attention to stocks of local companies than to those of nonlocal companies.

The literature provides conclusions that media information could be used as investor sentiment proxy. To explain the trading behaviors which cause the high volatility and trading volume in Bitcoin markets, scholars mainly measure investor sentiment and investor attention by different resource of outside internet data, which includes Google trend, twitter, and StockTwits. By deriving the information from Twitter and Google Trends, Philippos et al. (2019) find that Bitcoin prices are partially driven by a momentum on media attention. Shen et al. (2019) find that Tweets predict Bitcoin

returns, trading volume, and realized volatility. What's more, Ibikunle et al. (2020) argue that Google Trends drives the noise element of Bitcoin pricing. However, using messages on StockTwits, measuring investor sentiment, Guégan and Renault (2020) find that the significant relation between investor sentiment and Bitcoin returns disappears for lower frequencies of trading data.

### 2.3 Capturing investor sentiment from trading data

In addition to media information, literature has documented that New York Stock Exchange (NYSE) share turnover, the close-end mutual fund discount, the degree of underpricing in initial public offerings, and the aggregate level of corporate investment (Baker and Wurgler, 2006; Qui and Welch, 2004; Yu and Yuan, 2011). Also, daily trading data is used to proxy investor sentiment. Aboody et al. (2018) find that overnight returns could be used as firm-specific investor sentiment, but Xiong et al. (2020) argue that overnight return fails to serve as a proxy of firm-specific investor sentiment outside the U.S. stock market. By deriving from the interaction between financial market participants, Andersen et al. (2014) introduce a socio-financial model which capture the sentiment from the transition probabilities to change opinions (bullishness and bearishness) of the participants.

Both the inside markets data or the social-financial model employ different information resource to study investor sentiment in stock markets, which gives another method to investigate the influence of investor sentiment on Bitcoin pricing. The multiple methods of measuring investor sentiment in Bitcoin market help the research to have a more complete view.

## 3. Data and methodology

### 3.1 Data

Two sources of data are used in this paper. Firstly, Bitcoin data from Bitstamp<sup>1</sup>, the most popular and liquid bitcoin exchange in the US. We study the period from April 4<sup>th</sup>, 2014 to August 31, 2020. Secondly, we obtain investor sentiment data as Twitter data from <https://bitinfocharts.com/>, which captures the number of times the term 'Bitcoin' has been tweeted.

### 3.2 Methodology

To investigate the dynamics between investor sentiment, returns, realized volatility, trading volume, and liquidity, we estimate vector autoregressive (VAR) models. The variables selected in this paper are presented as follows:

1. Log-return:  $R_t = \log(P_t/P_{t-1})$ , where  $P_t$  is the daily close price of the Bitcoin series.
2. ILLIQ: Following Amihud (2002) and Choi (2020), the illiquidity is calculated as

---

<sup>1</sup> We download Bitstamp exchange tick data from [www.bitcoincharts.com](http://www.bitcoincharts.com).

$|R_t|/Volume_t$ , where  $Volume_t$  is the daily trading volume. the intuition is that price changes should be associated with large amounts of trading volumes to be liquid.

3. RV: Following Urquhart (2018), we calculate the daily realized volatility aggregating the data to the 5-min, which is  $RV_t = \sqrt{\sum_{j=1}^n r_{t,j}^2}$ , where  $r_{t,j}^2$  is the squared intraday 5-min interval log-returns and n is the number of the intervals.
4. Volume: Volume is the nature log of daily trading volume of the Bitcoin.
5. Bullishness: We use the socio-finance model to calculate the bullishness of the investors in Bitcoin, and a higher value of B indicates higher sentiment.
6. Tweet: Tweet is the nature log of the daily number of times the term “Bitcoin” has been tweeted.
7. GT: GT is the Google trend index of the term “Bitcoin”.

The VAR model estimated in this paper is as follows:

$$\chi_t = c + \sum_{j=1}^k \beta_j \chi_{t-j} + \varepsilon_t$$

where  $\chi_t$  is a vector that contains the variables of interest;  $c$  is a vector of constants;  $\varepsilon_t$  is a vector of independent white noise innovations. The lag length is determined by the Schwarz Bayesian information criterion and we estimate four separate models for each investor sentiment proxy to investigate whether the different investor sentiment can predict Bitcoin returns, liquidity, realized volatility and trading volume. Model 1 investigates the dynamics between Bitcoin returns and investor sentiment proxy, while Model 2 examine the relation between liquidity and investor sentiment proxy. Model 3 is used to capture the dynamics between realized volatility and investor sentiment proxy, and Model 4 for trading volume and investor sentiment proxy. We also employ Granger causality test for these models.

### 3.3 Simulation of the socio-finance model

To assess the performance of the inferential method based on MLE, we conduct a simulation study concerning the proposed socio-finance model. We set the sample size as 2000 ( $T = 2000$ ), and with four different parameter vectors. Different starting values of the model parameters are used to ensure that the maximization/minimization algorithm converges to the true simulated parameters. We simulate time series from the socio-finance model and estimate the parameters under the normality assumption.

Table 1 shows the first simulation with  $T = 2000$ , and using  $\lambda = 1.1$ ,  $\sigma_0 = 0.01$ ,  $\beta = 0.001$ , and  $\alpha = 400$  as the ‘true’ parameters. Column 1 of Table 1 shows the ‘true’ parameter vector  $\theta = (\lambda, \sigma_0, \beta, \alpha)'$  used to simulate the data, while column 2 shows the ‘true’ transformed parameter vector  $\theta^* = (\lambda^*, \sigma_0^*, \beta^*, \alpha^*)'$ . In column 3 we show the mean value of the 1000 starting-value vectors randomly generated. The maximum likelihood estimates (MLEs)  $\hat{\theta}^*$  of the transformed parameters  $\theta^*$  and the



corresponding standard errors are presented in column 4 and 5. Column 6 shows the value of the gradient of the log-likelihood function evaluated at the parameter estimates  $\widehat{\theta}^*$ . In the last column we show the estimated parameter vector  $\widehat{\theta}$ . We simulate 10 time-series in the first simulation, and estimate each of the time-series relying on 100 groups of starting values. Table 1 shows that mean value of  $\widehat{\lambda}^*$  (0.9421) deviates from its ‘true’ initial transformed value (0.0953) and is close to the mean value of the starting value (0.8664).  $\widehat{\sigma}_0^*, \widehat{\beta}^*, \widehat{\alpha}^*$  (-9.1343, -6.5993, 6.0610) are close to the ‘true’ initial transformed values (-9.2103, -6.9078, 5.9915), and independent from the starting values. The estimations of the first simulation give the evidence that the MLE is accurate for 3 of the 4 parameters of the socio-finance model. To further investigate the MLE, we have another 3 simulations with  $T = 2000$ .

**<Insert Table 1 Here>**

Table 2 shows the three simulations using three different ‘true’ initial parameter vectors, which are  $[2.8, 0.02, 0.04, 2.65]$ ,  $[4.2, 0.025, 0.05, 200]$ , and  $[12, 0.045, 0.07, 800]$ . We simulate 10 time-series using each of the three ‘true’ initial parameter vectors, and we estimate each of the time-series relying on 10 groups of starting values. For these three simulations, the transformed MLEs are close to the ‘true’ transformed initial values, and independent from the starting values. The four simulations and estimations give the evidence that the MLE method is suitable for estimating the parameters of the socio-finance model, that will be used in the empirical analysis in section 4 on the Bitcoin.

**<Insert Table 2 Here>**

The sample period of the Bitcoin data is from April 9, 2014 to August 31, 2020, 2337 trading days. We estimate the parameters using four different groups of starting values, and each group contains 10 different starting values vectors. To investigate the stability of the MLEs, we estimate the first period, second period, and the whole period respectively. The results in Table 3 shows that the estimates are stable across the period, which supports us to use the socio-finance model to capture the information of investor sentiment in Bitcoin market.

**<Insert Table 3 Here>**

#### 4. Empirical results

Table 4 shows the descriptive statistics for the logarithmic daily return (R), liquidity (ILLIQ), daily realized volatility (RV), the logarithmic daily trading volume (Volume), as well as the three different investor sentiment proxies, which are Bullishness (B), the logarithmic daily tweets (Tweet), and Google trend index (GT). The mean of the R is 0.0016 with the standard deviation being 0.0377 which are far larger, indicating that the daily return of Bitcoin is volatile. The values of the 25<sup>th</sup> and 75<sup>th</sup> provide the same

evidence. However, the daily realized volatility from the 5 min intervals data is relative stable with mean equaling 0.0374 and standard deviation 0.0264. We also find that the trading volume and liquidity are stable across the trading days in our sample. The statistics in our sample is similar to previous literature. The means of B, Tweet, and GT are -0.0079, 10.2269, and 32.3897. The B which is firstly introduced to measure Bitcoin investor sentiment shows that investors averagely hold bearish opinions.

**<Insert Table 4 Here>**

Table 5 shows the pairwise correlation coefficients of the variables in this paper. We find that the Bullishness is negative related to realized volatility and trading volume, and positive related to Tweet. The Tweet is positive related to the returns of Bitcoin. The G-trend is positive related to liquidity, realized volatility, and trading volume.

**< Insert Table 5 Here >**

Table 6 shows the results of the four VAR models where the estimated coefficients are showed in Panel A, and the results of Granger causality test are shown in Panel B. The results of Model 1 show that the autoregressive parameters for the Bullishness are significant for lag1. The results also indicate that past Bullishness have no significant impacts on future returns. Model 2 shows that the liquidity has significant autoregressive terms for lags 1, 2, 3, 4, 5 while Bullishness shows significant autoregressive terms for lag 3. The results also reveal that liquidity have significant influence on future Bullishness, which also supported by the Granger causality test. Being similar with Model 1 and 2, Model 3 shows that Bullishness have significant autoregressive terms. Moreover, Bullishness has significant influence on future realized volatility. This suggests that higher Bullishness is related to higher realized volatility in the future. This is also proven by the Granger causality test in Panel B. Model 4 investigates the interaction of trading volume and Bullishness and we find that the autoregressive Bullishness terms is significant at lags 1, 2, and 4. The Granger causality test also provides evidence that Bullishness Granger causes trading volume, which means higher Bullishness predicts higher future trading volume. Shen et al. (2019) employ twitter data to capture the information of well-informed investors and find that the tweets predict future realized volatility and trading volume, but not returns. Our Bullishness shows similar predictability as Shen et al. (2019), while we also prove that higher liquidity is related to lower future Bullishness.

**< Insert Table 6 Here >**

Table 7 shows the results of the four VAR models where the coefficients are presented in Panel A, and the Granger causality results are displayed in Panel B. In Model 1, we find that Tweet has significant autoregressive terms for lags 1, 4, 5, 6, and 7. The evidence also shows that the Bitcoin returns have significant impacts on future Tweet, which is not found in previous literature probably as we have longer sample. But from the perspective of Bullishness, that higher returns predict higher investor sentiment and attention is also proven by the results for Tweet. In model 2, liquidity shows significant autoregressive terms for lags 1 to 7 while Tweet also shows significant autoregressive terms for lags 1, 2, and 4 to 7. The results also indicate that higher levels

of tweets are associated with lower ILLIQ, which mean higher liquidity. Model 3 displays that significant autoregressive terms for both realized volatility (lags 1, 3, 4, 5) and Tweet (lags 1, 2, 4, 5, 6, 7). Moreover, we find that Tweet provides significant information about future realized volatility. Consistent with Shen et al. (2019), we also find significant autoregressive terms for both trading volume (lags 1, 3, 4, 6, 7, 8) and Tweet (lags 1, 2, 4, 5, 6, 7). Using a longer sample, we also find Tweet can predict future trading volume, saying more tweets more trades. The relations between Bitcoin return, liquidity, realized volatility, trading volume and Tweet investigated in Panel A are also proven in Panel B by Granger causality tests.

< **Insert Table 7 Here** >

As one of the most important two investor information sources of Bitcoin, Google trend is also investigated in this paper. Table 8 shows the analyses of the relation between Bitcoin return, liquidity, realized volatility, trading volume and Google trend. Model 1 shows that Google trend has significant autoregressive terms for lags 1, 2, and 4 to 7. No significant results indicate predictability between Bitcoin return and Google trend. In Model 2, we find significant estimates of the autoregressive parameters for the liquidity for lags 1, 3, and 5 to 8 while Google trend also show significant autoregressive terms for lags 1, 2, 4, 5, 6, and 8. Combining the Granger causality test in Panel B, we find higher Google trend index indicates higher future liquidity. Model 3 reports the interaction between realized volatility and Google trend. The realized volatility has significant autoregressive terms for lags 1, 3, 4, and 6 while the Google trend has significant autoregressive terms for 1, 2, 4, and 7. We also find realized volatility capture significant information of future Google trend, which is consistent with Urquhart (2018) and proven by Granger causality test. Model 4 shows significant autoregressive terms for trading volume (lags 1 to 4 and 6 to 8) and Google trend (lags 1, 2, 4, 8). Different from Urquhart (2018), we find Google trend could predict future trading volume, which means higher search queries will cause more trades of Bitcoin. The Granger causality test supports the relation.

< **Insert Table 8 Here** >

## 5. Conclusion

By introducing a new investor sentiment proxy in Bitcoin, which is Bullishness, we find that Bullishness capture information of informed investors. Similar to Tweet, Bullishness is significant related to future realized volatility and trading volume. Different from Google trend, which is thought to measure uninformed individuals, Bitcoin returns have significant influence on future Bullishness and Tweet. We also find that only tweet has significant impacts on liquidity.

Relying on our longer sample, we find that Google trend has significant influence on future trading volume, which means that the investors in Bitcoin are becoming more informed to some extent. Thus, the evidence supports that Bullishness can serve as a well-informed investor sentiment proxy.

## Appendix. Inference of the socio-finance model

### A1. the socio-finance model

Relying on the diffusion and influence between investors' opinions (bullishness or bearishness) on the assets, Andersen et al. (2014) define the socio-finance model by a quantitative manner to exactly describe the market-wide opinions on a certain asset. Figure A1 shows the changing bullishness in a population via communications in subgroups.

**<Insert Figure A1 Here>**

The participants in the market are shown schematically as circles in Figure A1A. Following the Galam model (Galam, 2005; Galam, 2012; Biondi et al., 2012), we proceed the opinion formation by simplicity imagine that people have just two mutually exclusive opinions on the asset, which are characterized as either 'bullish' (black circles) or 'bearish' (white circles) in Figure A1. Figure A1A presents the opinions of the participants in the market at the beginning of a given day. We assume that the participants meet in subgroups of different sizes, which are illustrated by the different boxes in Figure A1B, to update their opinions the asset. For example, there are six persons in the leftmost box in Figure A1B with two bullish and four bearish, whom can be imagined as sitting at a table, or having an online conference, discussing their opinion on the latest market developments. Figure A1C illustrates the outcome of the discussions for the different groups. After one round communication, Figure A1D shows the population bullishness. We introduce majority principle to the opinion formation in each subgroup. The majority opinion in a given group manages to polarize the opinion of the group by changing the opinion of minority. If we take the afore mentioned group of six persons we can see that after discussing, because of the majority principle, they have all become bearish. Considering more realistically, we instead assume that is a certain probability for a majority opinion to prevail, and that even under certain conditions a minority could persuade a part of the majority to change their opinions.

To quantitative describe the bullishness opinion formation progress, we define that a given group of size  $k$  with  $j$  agents having a bullish opinion and  $k-j$  a bearish opinion and let  $m_{k,j}$  denotes the transition probability for all the  $k$  agents to adopt the bullish opinion after the meeting. After one update taking into account communications in all groups of size  $k$  with  $j$  bullish agents, the updated probability of finding an agent with a bullish opinion in the population can therefore be written as following:

$$B(t+1) = m_{k,j}(t) C_k^j B(t)^j [1 - B(t)]^{k-j} \quad (1)$$

$$C_k^j \equiv \frac{k!}{j! (k-j)!}$$

where  $B$  is the proportion of the population holding bullishness opinions.  $C_k^j$  is the binomial coefficients that indicates the arrangement of the group with size  $k$  has  $j$  agents

holding bullishness opinions.  $m_{k,j}$  represents the transition probability that an agent changing his/her bullishness opinion to bearish. We assume that the transition probability  $m_{k,j}$  changes as the market performance changes, depending on time.

Taking the sum over different groups of different sizes and different composition of bullishness within each group, we obtain the general  $B(t + 1)$ , for the bullishness in a population at time  $t+1$  due to the outcome of meetings at time  $t$ :

$$B(t + 1) = \sum_{k=1}^L a_k \sum_{j=0}^k m_{k,j}(t) C_j^k B(t)^j [1 - B(t)]^{k-j} \quad (2)$$

$$\sum_{k=1}^L a_k = 1, a_k \equiv \frac{1}{L}$$

with  $L$  denoting the size of the largest group and  $a_k$  denoting the weight of the group of size  $k$ . The link between communication and its impact on the markets can be taken into account by assuming that the return series  $r$  changes whenever there is a change in the bullishness. The idea is that the bullishness itself is not the relevant factor determining how prices will change. Investors feeling bullish would naturally already hold long positions on the asset. Rather, when people change their opinion, say becoming more negative about the market, or less bullish, this will increase their tendency to sell. The absolute sentiment can act as a contrarian indicator for future asset returns. We assume the return to be proportional to the percentage change in bullishness,  $RB(t)$ , as well as the economic news,  $\eta(t)$ , the return  $r$  is given by:

$$r(t) = \frac{1}{\lambda} RB(t) + \eta(t), \lambda > 0 \quad (3)$$

where  $RB(t) = \frac{B(t) - B(t-1)}{B(t-1)}$  is the change of the bullishness.  $\lambda$  is the parameter that connects the change of the bullishness and returns. The  $\eta(t)$  is assumed to be either Gaussian or Student-t distributed with mean zero and a standard deviation that varies as a function of time depending on the changes in sentiment. We assume that the market react to fundamental economic news represented by  $\eta$  but that the amplitude of the reaction depends on changes in the sentiment  $RB$ :

$$\sigma(t) = \sigma_0 \exp\left[\frac{1}{\beta} |RB(t)|\right], \sigma_0 > 0, \beta > 0 \quad (4)$$

where the  $\sigma$  is the deviation of the  $\eta$ ;  $\sigma_0$  is the parameter represent the initial deviation of the  $\eta$ ;  $\beta$  is the parameter that represents how the  $RB$  influences the  $\sigma$ .

The influence of the financial market performance on decision-making can now be included in a natural way by letting the strength of persuasion depend on how the market has performed since the last meeting of the market participants. For example, if the market had a dramatic downturn at the close yesterday, then in the meetings the next morning, investors with a bearish opinion will be more likely to convince others even in

a bullish majority group. We take into account the transition probability of the opinions which depends on the market return over the last period:

$$m_{k,j}(t) = m_{k,j}(t-1) \exp\left[\frac{1}{\alpha} r(t)\right]; m_{k,j}(t=0) \equiv j/k, \alpha > 0 \quad (5)$$

where  $\alpha$  defines the scale for which a given return  $r$  impacts the transition probabilities. The condition  $m_{k,j}(t=0) \equiv j/k$  describes the initially unbiased case where in average no market participant changes opinion.

## A2. Estimation of the socio-finance model

We present the inferential method adopted to estimate the parameters of the socio-finance model. The method on which estimation will be based is *maximum likelihood*. Let  $\theta$  denotes the parameter vector to be estimated and  $r = (r(1), r(2), \dots, r(T))$  is the return sample with size  $T$ . We calculate the joint probability density:

$$f_{R(1),R(2),\dots,R(T)}(r(1), r(2), \dots, r(T)|\theta) \quad (6)$$

which can be viewed as the probability of having observed this particular sample<sup>2</sup>. The maximum likelihood estimate of  $\theta$  is the vector  $\hat{\theta}$  that maximizes eq. (7). To find maximum likelihood estimates, we firstly calculate the likelihood function and to find values of  $\theta$  that maximize the function.

Under the assumption of a normal and a Student-t distribution for the error process  $\eta(t)$ , we calculate the likelihood function, and maximize the function based on numerical optimization algorithms. The likelihood function i.e. the joint probability density of the return sample  $r = (r(1), r(2), \dots, r(T))$  for the proposed socio-finance model is:

$$\begin{aligned} & f[r|B(0), m_{k,j}(0), \theta] \\ &= f[r(T)|r(1), \dots, r(T-1), B(0), m_{k,j}(0), \theta] \\ & \cdot f[r(1), \dots, r(T-1)|B(0), m_{k,j}(0), \theta] \\ &= f[r(T)|r(1), \dots, r(T-1), B(0), m_{k,j}(0), \theta] \\ & \cdot f[r(T-1)|r(1), \dots, r(T-2), B(0), m_{k,j}(0), \theta] \\ & \cdot f[r(1), \dots, r(T-2)|B(0), m_{k,j}(0), \theta] = \dots \\ &= f[r(1)|B(0), m_{k,j}(0), \theta] \cdot \prod_{t=2}^T f[r(t)|\Phi(t-1), B(0), m_{k,j}(0), \theta] \end{aligned} \quad (7)$$

where  $\Phi(t-1)$  is the information set up to time  $t-1$ .

<sup>2</sup> With capital letters,  $R(t)$ , we denote the random variable of the return at time  $t$ , while with small letters,  $r(t)$ , we denote a particular value that the random variable takes at time  $t$ .

Based on this property, that the joint probability can be written as a product of conditional probabilities, we calculate the likelihood function as follows. First, we consider the probability distribution of  $r(1)$  under the normality assumption for the error process  $\eta(t)$ .  $r(1)$  is Gaussian and the density, conditional on  $B(0)$  and  $m_{k,j}(0) = j/k$ , taking the form:

$$f[r(1)|B(0), m_{k,j}(0), \theta_N] = \frac{1}{\sqrt{2\pi}\sqrt{\sigma(1)^2}} \exp \left\{ -\frac{1}{2\sigma(1)^2} \left[ r(1) - \frac{1}{\lambda} RB(1) \right]^2 \right\}$$

where  $\theta_N = (\lambda, \sigma_0, \beta, \alpha)'$  denotes the parameter vector to be estimated under the normality assumption.

Conditioning on  $r(1)$ , the density of the second observation  $r(2)$  is:

$$f[r(2)|r(1), B(0), m_{k,j}(0), \theta_N] = \frac{1}{\sqrt{2\pi}\sqrt{\sigma(2)^2}} \exp \left\{ -\frac{1}{2\sigma(2)^2} \left[ r(2) - \frac{1}{\lambda} RB(2) \right]^2 \right\}$$

Proceeding in this fashion, the conditional density of the  $t$ -th observation can be calculated as:

$$\begin{aligned} f[r(t)|r(t-1), \dots, r(1), B(0), m_{k,j}(0), \theta_N] \\ = \frac{1}{\sqrt{2\pi}\sqrt{\sigma(t)^2}} \exp \left\{ -\frac{1}{2\sigma(t)^2} \left[ r(t) - \frac{1}{\lambda} RB(t) \right]^2 \right\} \end{aligned} \quad (8)$$

Therefore, the likelihood of the complete sample is:

$$\begin{aligned} f[r|B(0), m_{k,j}(0), \theta_N] \\ = (2\pi)^{-\frac{T}{2}} \prod_{t=1}^T [\sigma(t)^2]^{-1/2} \exp \left\{ -\frac{1}{2} \sum_{t=1}^T \frac{1}{\sigma(t)^2} \left[ r(t) - \frac{1}{\lambda} RB(t) \right]^2 \right\} \end{aligned} \quad (9)$$

The log-likelihood function, denoted  $L_N(r|\theta_N)$ , is:

$$L_N(r|\theta_N) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T [\ln \sigma(t)^2] - \frac{1}{2} \sum_{t=1}^T \frac{1}{\sigma(t)^2} \left[ r(t) - \frac{1}{\lambda} RB(t) \right]^2 \quad (10)$$

The vector of  $\theta_N$  that maximizes the conditional likelihood function (8) is identical to the vector that maximizes the conditional log-likelihood function (9).

Although the normal distribution is the most commonly used in applications, there is empirical evidence that the distribution of financial time series has usually fat tails, even after taking into account the volatility clustering phenomenon. In other words, the normality assumption of standardized residuals of estimated financial models is usually

rejected in most financial applications. A solution to this problem is to specify a distribution that accounts for fat tails and deviations from normality such as the Student-t (Bollerslev, 1987) or a Generalized Error distribution (Nelson, 1991). Under the Student-t distribution with  $\nu$  degrees of freedom for the error process  $\eta(t)$ , the likelihood for the socio-finance model for the complete sample can be written as:

$$\begin{aligned}
& f_{ST}[r|B(0), m_{k,j}(0), \boldsymbol{\theta}_{ST}] \\
&= \left[ \Gamma\left(\frac{\nu+1}{2}\right) \right]^T \left[ \Gamma\left(\frac{\nu}{2}\right) \right]^{-T} [\pi(\nu-2)]^{-T/2} \prod_{t=1}^T [\sigma(t)^2]^{-1/2} \prod_{t=1}^T \left[ 1 + \frac{(r(t) - \frac{1}{\lambda}RB(t))^2}{(\nu-2)\sigma(t)^2} \right]^{-(\nu+1)/2}
\end{aligned} \tag{11}$$

where  $\Gamma(\cdot)$  is the gamma function. The log-likelihood function is:

$$\begin{aligned}
& L_{ST}(r|\boldsymbol{\theta}_{ST}) \\
&= T \ln \Gamma\left(\frac{\nu+1}{2}\right) - T \ln \Gamma\left(\frac{\nu}{2}\right) - \frac{T}{2} \ln[\pi(\nu-2)] \\
&\quad - \frac{1}{2} \sum_{t=1}^T [\ln \sigma(t)^2] - \frac{\nu+1}{2} \sum_{t=1}^T \ln \left[ 1 + \frac{(r(t) - \frac{1}{\lambda}RB(t))^2}{(\nu-2)\sigma(t)^2} \right]
\end{aligned} \tag{12}$$

where  $\boldsymbol{\theta}_{ST} = (\lambda, \sigma_0, \beta, \alpha, \nu)'$  denotes the parameter vector to be estimated under the Student-t distribution.

To avoid the positivity restrictions for the parameters,  $\lambda > 0$ ,  $\sigma_0 > 0$ ,  $\beta > 0$ , and  $\alpha > 0$ , we use the logarithmic transformation, so that  $\lambda^* = \ln(\lambda)$ ,  $\sigma_0^* = \ln(\sigma_0^2)$ ,  $\beta^* = \ln(\beta)$ , and  $\alpha^* = \ln(\alpha)$ . For the degrees of freedom parameter  $\nu$ , we use the transformation  $\nu^* = \ln(\nu-2)$ . The parameter vector to estimate is  $\boldsymbol{\theta}_N^* = (\lambda^*, \sigma_0^*, \beta^*, \alpha^*)'$  under conditional normality, and  $\boldsymbol{\theta}_{ST}^* = (\lambda^*, \sigma_0^*, \beta^*, \alpha^*, \nu^*)'$  under the Student-t error distribution. We reparametrize the parameters to ensure that a numerical optimization algorithm always provides parameter values within certain specified boundaries. The maximum likelihood solution is invariant under transformation of parameters to log-likelihood. We use numerical optimization algorithms to maximize the conditional log-likelihood function, to deal with the highly non-linear nature of the proposed socio-finance model. In this paper, we maximize the log-likelihood function by applying the optimization function '*fminunc*' of Matlab.



## Reference

- Andersen, J. V., I. Vrontos, P. Dellaportas, and S. Galam. (2014). Communication impacting financial markets. *EPL (Europhysics Letters)*, 108 (2): 28007.
- Aboody, D., Even-Tov, O., Lehavy, R., Trueman, B. (2018). Overnight returns and firm-specific investor sentiment. *Journal of Financial and Quantitative Analysis*, 53, 485-505.
- Antweiler, W., Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance*, 59, 1259-1294.
- Baker, M., Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, 61, 1645-1680.
- Baker, M., and J. Wurgler. (2007). Investor Sentiment in the Stock Market. *The Journal of Economic Perspectives*, 21 (2): 129-151.
- Barber, B. M., Odean, T. (2008). All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, 21, 785-818.
- Barber, B. M., Odean, T., Zhu, N. (2009). Do retail trades move markets? *Review of Financial Studies*, 22, 151-186.
- Bollen, J., Mao, H., Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2, 1-8.
- Choi, H. (2020). Investor attention and bitcoin liquidity: Evidence from bitcoin tweets. *Finance Research Letters*:101555.
- Da, Z., Engelberg, J., Gao, P. (2015). The sum of All FEARS investor sentiment and asset prices. *Review of Financial Studies*, 28, 1-32.
- Dimpfl, T., Jank, S. (2016). Can internet search queries help to predict stock market volatility? *European Financial Management*, 22, 171-192.
- Eom, C., T. Kaizoji, S. H. Kang, and L. Pichl. (2019). Bitcoin and investor sentiment: Statistical characteristics and predictability. *Physica A: Statistical Mechanics and its Applications*, 514: 511-521.
- Galam, S. (2012). *A Physicist's Modeling of Psycho-political*. Springer, New York.
- Galam, S. (2005). Local dynamics vs. social mechanisms: A unifying frame. *Europhysics Letters (EPL)*, 70 (6): 705-711.
- Gao, B., and C. Yang. (2018). Investor Trading Behavior and Sentiment in Futures Markets. *Emerging Markets Finance and Trade*, 54 (3):707-720.
- Guégan, D., and T. Renault. (2020). Does investor sentiment on social media provide robust information for Bitcoin returns predictability? *Finance Research Letters*: 101494.
- Hillert, A., Jacobs, H., and Müller, S. (2014). Media makes momentum. *Review of Financial Studies*, 27, 3467-3501.
- Huang, Y., Qiu, H., Wu, Z. (2016). Local bias in investor attention: evidence from

- China's internet stock message boards. *Journal of Empirical Finance*, 38, 338-354.
- Huang, D., F. Wen, and X. Yang. (2009). Investor sentiment index and empirical evidence from China's stock market. *Journal of Systems Science & Mathematical Sciences*, 29(1), 1-13.
- Joseph, K., Babajide Wintoki, M., Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: evidence from online search. *International Journal of Forecasting*, 27, 1116-1127.
- Kim, S.-H., Kim, D. (2014). Investor sentiment from internet message postings and the predictability of stock returns. *Journal of Economic Behavior & Organization*, 107, 708-729.
- KUMAR, A., and C. M. C. LEE. (2006). Retail Investor Sentiment and Return Comovements. *The Journal of Finance*, 61 (5):2451-2486.
- Leung, H., Ton, T. (2015). The impact of internet stock message boards on cross-sectional returns of small-capitalization stocks. *Journal of Banking & Finance*, 55, 37-55.
- Li, X., Shen, D., Zhang, W. (2018). Do Chinese internet stock message boards convey firm-specific information? *Pacific-Basin Finance Journal*, 49, 1-14.
- Li, C.-A., and C.-C. Yeh. (2011). Investor Psychological and Behavioral Bias: Do High Sentiment and Momentum Exist in the China Stock Market? *Review of Pacific Basin Financial Markets and Policies*, 14 (03):429-448.
- Ni, Z.-X., D.-Z. Wang, and W.-J. Xue. (2015). Investor sentiment and its nonlinear effect on stock returns—New evidence from the Chinese stock market based on panel quantile regression model. *Economic Modelling*, 50:266-274.
- Philippas, D., H. Rjiba, K. Guesmi, and S. Goutte. (2019). Media attention and Bitcoin prices. *Finance Research Letters*, 30: 37-43.
- Qiu, L., and I. Welch. (2006). Investor sentiment measures: National Bureau of Economic Research.
- Ryu, D., H. Kim, and H. Yang. 2017. Investor sentiment, trading behavior and stock returns. *Applied Economics Letters*, 24 (12):826-830.
- Shen, D., A. Urquhart, and P. Wang. (2019). Does twitter predict Bitcoin? *Economics Letters*, 174:118-122.
- Siganos, A., Vagenas-Nanos, E., Verwijmeren, P. (2014). Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior & Organization*, 107, Part B, 730-743.
- Solomon, D.H. (2012). Selective Publicity and Stock Prices. *Journal of Finance*, 67, 599–637.
- Sprenger, T. O., Tumasjan, A., Sandner, P. G., Welpe, I. M. (2014). Tweets and Trades: the Information Content of Stock Microblogs. *European Financial Management*, 20(5),

926-957.

Tumarkin, R., Whitelaw, R. F. (2001). News or noise? Internet postings and stock prices. *Financial Analysts Journal*, 57, 41-51.

Urquhart, A. (2018). What causes the attention of Bitcoin? *Economics Letters*, 166:40-44.

Wysocki, P. (1998). Cheap talk on the web: The determinants of postings on stock message boards. University of Michigan Business School Working Paper (98025).

Xiong, X., Y. Meng, X. Li, and D. Shen. (2020). Can overnight return really serve as a proxy for firm-specific investor sentiment? Cross-country evidence. *Journal of International Financial Markets, Institutions and Money*, 64:101173.

Yu, J., and Y. Yuan. (2011). Investor sentiment and the mean–variance relation. *Journal of Financial Economics*, 100 (2):367-381.

Zhang, Y., Song, W., Shen, D., & Zhang, W. (2016). Market reaction to internet news: Information diffusion and price pressure. *Economic Modelling*, 56, 43-49.