Coin Market Behavior using Social Sentiment Markov Chains

Completed Research Paper

Kwansoo Kim

Sang-Yong Tom Lee

Said ASSAR

Abstract

2017 was the pivotal year for the cryptocurrency made its foray into the mainstream financial markets. However, in 2018 it was a panic itself. For the market, it is imperative to ruminate reasonable analysis and expectations compiling various situations. We examine the dynamic interactions between the coin market behavior and market events and social metrics to incorporate potential coin traders' social belief and response into the market moves. We use hidden Markov model performing four different experiments in two-coin markets. We show that market events such as closing price and trading volume are important predictors for the coin market behavior, but not equally in all experiments. Interestingly, hidden Markov process demonstrates that social media metrics fairly explain the future behavior of the coin markets. Overall, this study shows the combined impact of market events and social factors through varied sequences adjusted for each market situation on the coin market behavior.

Keywords: Cryptocurrency, coin market behavior, coin investor, market events, social sentiment, Markov chains, text analysis

Introduction

Cryptocurrency derived from the block chain technology, such as bitcoin, has received considerable attention in recent years and is expected to affect the overall economy. Tapscott (2016) noted that the cryptocurrency would exert a radical impact on the financial system. To date, bitcoin is the most significant example of blockchain-based cryptocurrencies (Li and Wang 2017).

Years back, the European Central Bank (2012) situated bitcoin as unregulated digital money, a virtual currency, which in contrast to paper money and can be used on the Internet (Zähres 2012). Bitcoin does not represent a banknote such as dollars and has its own value units. In this sense, the coin is resemblant to other alternative currencies (Polasik et al. 2016). However, it is significantly different from local currencies, which are equivalent to the official currency, guaranteed by the issuers. In addition, by using a decentralized leger system of blockchain, bitcoin technically differs from virtual currencies (Jin and Bolebruch 2009).

Digital currency was first emerged in the 1990s for peer to peer (p2p) payments (Clemons 1996). In conjunction with this, bitcoin is defined as a new digital currency using cryptography and information technology to accelerate P2P transactions (Mai et al. 2018). In the capital market, Hileman (2016) estimated that 12 million trading accounts and over 100,000 retailers worked with bitcoin in the fourth quarter of 2015. Böhme et al. (2015) showed that bitcoin has appeared to be a fintech innovation, disrupting existing payments and monetary systems in a short period.

Based on the work of Delone and Mclean (2004), we imply that the increase in value and high rate of return given to participants became involved in the development of bitcoin could be regarded as a measure of information system success. This denotes that bitcoin has the potential to make an impact on the financial technology industry. Thus, identifying what determines bitcoin's value and finding the factors that influence its monetary value has become an important issue (Mai et al. 2018).

However, this study is laying ever more stress on predicting coin market movements than estimating the cryptocurrency value. To this end, we examine the overall situation of the coin markets and place an emphasis on the defined research objective.

If we look at the coin markets in Korea, in 2017 it was a good time to make money by purchasing bitcoin without knowing the basic idea of a decentralized distribution ledger using a hash function and a public key. It was not a bitcoin though. Coin traders could earn money by investing any other cryptocurrencies. Whenever a new coin was listed on the exchange, they would not ask, did buy it. Nonetheless, those coin holders were able to sell it with a profit of dozens of times.

However, the situation has changed. According to the coin market announcement, bitcoin price peaked, and then plunged 40% at the beginning of 2018. The downturn was moderated in the second quarter, and the coin price stagnated in the third quarter, but dropped sharply in the last quarter. Conclusionally**,** its value was not less than one fifth of a year ago at the end of 2018. In addition, Altcoins[1] dropped more significantly than bitcoin, and most suffered losses of more than 90%.

In 2018, many investors who ran into the coin markets late could not earn a speculative gain. Most of them sold off all the coins at a large loss due to price drop. For the past two years, most investors have traded without knowing the driver of change in the coin price. They just followed the bitcoin price trend. Or they only bought and sold bitcoin in line with fluctuations in the price. Thus, it was significant to analyze the factors to identify the cause of the repetitive and persistent decline of bitcoin price in 2018. Related experts denote that the loss of faith in coin markets is the main factor due to government regulations and the lack of initial coin offering[2] (ICO) project success. The price trend

---

[1] Altcoins are the alternative cryptocurrencies launched after the success of bitcoin. Generally, they project themselves as better substitutes to bitcoin. The success of bitcoin as the first peer-to-peer digital currency paved the way for many to follow. Many altcoins are trying to target any perceived limitations that bitcoin has and come up with newer versions with competitive advantages. As the term 'altcoins' means all cryptocurrencies which bitcoin are not. https://www.investopedia.com/terms/a/altcoin.as

[2] A company can start small and grow as its profits allow, remaining beholden only to company owners but having to wait for funds to build up. Alternately, companies can look to outside investors for early support, providing them a quick influx of

was more influential by the external factor e.g., what profit strategy was established by the financial forces in the markets, than the internal factor in the markets.

Under the circumstance, studying the coin market behavior has important practical and theoretical implications. Coin investors can foresee the movements of the price and trading volume to estimate the expected return. Policymakers can regulate the market forces behind the cryptocurrency to secure financial stability (Japson 2016). Financial business can understand the varying patterns in price and trading volume before adopting the coin or even launching their own digital currency (Ren and Culpan 2017). Information systems researchers need to understand how the combination of market factors and social media affect the blockchain driven coin market in order to advance IS theory by identifying the roles of different parties in the dispersion of new financial technology.

Now, we are getting aware of the importance of coin market-related information to be expressed in social media. For example, an information service organization in Japan has announced a trial launch of a service that distributes information related to cryptocurrency collected from SNS using artificial intelligence (AI). The service will foster an environment in which coin investors will be able to trade with confidence by providing the important and reliable information about cryptocurrency without delay. The information service is developing a system to evaluate the importance of tweets, which will be culled from Twitter. The information gathered is going to be distributed through cloud platform. The service will pass around nuance analysis of not only news that may affect the coin markets, but also positive and negative perceptions of market data such as buzzwords. This is a good example to represent a close relationship between the coin market and social media.

The coin market has similar features to those of the stock market. Therefore, prior studies explored the coin market using models for stock markets. Glaser et al. (2014) examined coin holders' motivation and gauged that the investors treat their coin as an asset rather than payment. Kritoufek (2013, 2015) detected that bitcoin has been linked to conventional google search. The coin's popularity on the search engine are highly associated with bitcoin exchange rates (80.6 percent) and weekly total transaction volume (89.1 percent) at the top four exchanges. The tight grip between online users' interest and bitcoin valuation is like the connection between web visits and firm equity value (Dewan et al. 2002).

Relevantly, prior studies discussed economic issues in cryptocurrency systems e.g., the economic function, mechanism, and value of cryptocurrencies (Bohme et al. 2015, Evans 2014, Yermack 2013). The results revealed that the exchange trade ratio and speculative behavior play a significant role in lower frequencies and that the Chinese market index may be a main driver of bitcoin price. Garcia (2014) examined the impact of online word of mouth in social platforms on top of search. Bouoiyour and Selmi (2015) identified a set of indicators, including google search, ratio of exchange trade volume, the hash rate, and stock market. Ciaian et al. (2016) found that transaction volume, user volume, and attractiveness have significant impacts on bitcoin price. They also found significant changes over time. Polasik et al. (2015) studied the impact of news articles volume, sentiment, google search, transaction amount, number of bitcoins, and economic factors such as industrial production growth, unemployment, and inflation on the monthly return of bitcoin and found that returns are driven primarily by news volume, news sentiment, and the total number of transactions.

In this context, we examine the research questions. Is there a predictive relationship between social trend and sentiment, and coin market behavior? How are the social factors effective over market events to predict the coin market behavior? Does it yield the same results in all coin markets?

We assembled various data from the coin markets, conventional stock markets, and social media. We performed sentiment analyses of tweets after utilizing Korean morphological analyzer. Basically, we used kernel regularized least square to empirically test the relationship between the coin market behaviors and social factors. However, to identify a more accurate relationship, we employed the Markov chains process, which has been used in this context. We modeled returns as a two hidden states Markov chains, so that daily return is designated as either state upward or downward. Markov

---

cash but typically coming with the trade-off of giving away a portion of ownership stake. https://www.investopedia.com/terms/i/initial-coin-offering-ico.asp chains allows for nonlinearity of the return series e.g., we can test if a high return is more likely after a few days of low returns than a few days of high returns. This method does not demand a normal distribution of returns even though the returns need to be stationary for constant transition probabilities.

Our findings illustrate that market events are important predictors of the coin market behaviors and prove the overall situation of the market previously described. That is, coin investors only bought and sold bitcoin in line with fluctuations in the coin price without knowing the driver of change in the price. However, the results also demonstrate that social factors influence the coin market behavior, but not equally in all coin markets. Hidden Markov model suggests that social media metrics fairly explain the future behavior of the coin markets. One experiment shows that social media metrics appeared to be more effective than previous market events in predicting coin market behavior.

This study has two implications. We upbuild a comprehensive understanding of the social factors behind the coin market behavior. We show that social media is a potential significant source that can explain and predict the coin market. Our findings provide a different perspective on the market prediction and the fintech proliferation compared to the prior studies identifying what determines bitcoin's value. In addition, we contribute to IS theory by highlighting the different influences of diverse compounds of market events and social factors. We experiment four different cases in two-coin markets. We show that internal and external market events and social factors play an important role of impacting the market behavior in each case, but not all are influential equal.

Our paper is organized as follows. Section 2 of the paper presents data and variables, which were created for research objectives. Section 3 describe the model development using a procedure based on the use of Markov chains. Section 4 provide the findings. Section 5 summarizes the results and discusses the implications and conclusions.

Data and Variables

All data were collected daily. The data collection period was from November 2017 to April 2018. Google trend was used to collect web search traffic data that indicates how many times a specific keyword was searched in Google. It enables us to easily identify users' search term trend. Data in each point is divided by the total searches of the geography and time range in order to compare relative popularity. That is, the resulting numbers then get scaled on a range of 0 to 100 based on a topics proportion to all searches. We set the region to Korea.

Social sentiment data was collected by crawling bitcoin related posts on Twitter. The social platform provides API enabling people to easily crawl objects e.g., posts, users, places, etc. But it does not allow us to access large amount of data or past data before a certain period. Thus, we constructed web crawler using Selenium and BeautifulSoup library of Python for the purpose.

The data set is that a total of 154,783 tweets were crawled. However, the data included noises such as advertisements. Thus, we secured a total of 100,120 data after filtering. And then, we preprocessed the data by removing special characters and converted the data into the proper form for morphological analysis. Finally, sentiment analysis was applied to the transformed data to classify them as positive and negative.

In Korea, we do not have any officially licensed market for cryptocurrency, unlike other financial products, and do not have any integrated index which shows trends and flows of fund such as NASDAQ. Thus, coin investors use exchanges to trade the cryptocurrency. However, the coin type, price and trading volume are different for each exchange. We selected two exchanges, *Bithumb* and *Upbit*, most used by coin traders in order to collect data of bitcoin price and trading volume. The coin markets are open for 24 hours. Thus, we used the closing price and the daily total trading volume provided by each exchange. Table 1 describes variables we used for the analysis.

Table 1. Variable and Definition

Variable	Operational Definition
i	Cryptocurrency market (i: Upbit or Bithumb)
t	Calendar time in days (1, 181)
Coin market	
Trading volume for bitcoin_{it}	Trading volume of bitcoin in market i at day t
Closing price for bitcoin_{it}	Closing price of bitcoin in market i at day t
Stock market	
Trading volume_t	Trading volume of stock market at day t
Closing price_t	Closing price of stock market at day t
Social Media Sentiment	
Google trend_t	Number of bitcoin searched on Google trend at day t
Positive sentiment_t	Number of positive sentiment about bitcoin on Twitter at day t
Negative sentiment_t	Number of negative sentiment about bitcoin on Twitter at day t

However, we want to examine how the coin market is influenced by pervious market events and social factors. We study how the coin markets behave with those factors and predict how the markets move. To this end, we use a percentage change to express a change in a variable, which represents the relative change between values of two time points. The coin market movements can be expressed in terms of trading volume and price directions. Thus, we created two percentage changes, *rate is (present value or volume – old value or volume) / (old value or volume) and direction is (new value or volume – present value or volume) / (present value or volume)* as followings.

Table 2. Variable Creation

Variable	Operational Definition
Coin market	
Trading volume rate	(trading volume – lagged trading volume) / lagged trading volume
Closing price rate	(closing price – lagged closing price) / lagged closing price
Trading volume direction	(leading trading volume – trading volume) / trading volume
Closing price direction	(leading closing price – closing price) / closing price
Stock market	
Market trading volume rate	(market volume – lagged market volume) / lagged market volume
Market price rate	(market price volume – lagged market price) / lagged market price
Social Media Sentiment	
Google trend rate	(google trend – lagged google trend) / lagged google trend
Positive sentiment rate	(positive sentiment – lagged positive sentiment) / lagged positive sentiment
Negative sentiment rate	(negative sentiment – lagged negative sentiment) / lagged negative sentiment
Note. we used 181 days' time period. rate= (day_t – day_{t-1})/ day_{t-1}, direction = (day_{t+1} – day_t)/ day_t	

Research Model Development

First, we employed Kernel Regularized Least Squares (KRLS) to identify the impact of those factors on the coin market behavior. The method mitigates a conventional assumption that the marginal effect of the explanatory factor is constant and reduces misspecification bias compare to Ordinary Least Squares (OLS). In addition, KRLS uses regularization to minimize overfitting problem and to diminish the influence of outliers (Haimmueller and Hazlett 2014).

Table 3 depicts the impact of pervious market events and social factors on coin market behaviors. Our interest variables are social factors. The results show that only positive sentiment rate (social factor) statistically significantly increases the trading volume in *Bithumb*. On the other hand, trading volume rate (a market event) has a negative effect on the coin market behavior across two markets in a rough way. At this point, we do not clarify if social factors influence the coin market behavior.

Table 3. KRLS Estimation Results

Variable	Closing Price Direction (Bithumb)	Trading Volume Direction (Bithumb)	Closing Price Direction (Upbit)	Trading Volume Direction (Upbit)
Google trend rate	-0.00049 (0.00107)	0.07786 (0.06941)	-0.00026 (0.00103)	0.03832 (0.04485)
Positive sentiment rate	0.00053 (0.00054)	0.08630** (0.04226)	-0.00015 (0.00052)	0.00250 (0.02463)
Negative sentiment rate	0.00022 (0.00052)	0.00098 (0.03774)	-0.00016 (0.00049)	0.01768 (0.02294)
Trading volume rate	-0.00113 (0.00077)	-0.16375*** (0.04994)	-0.00146** (0.00075)	-0.10412*** (0.03166)
Closing price rate	0.00493 (0.00494)	-0.73750** (0.32241)	0.00501 (0.00473)	-0.19328 (0.20810)
Market volume rate	0.00091 (0.00203)	0.25578** (0.11938)	0.00043 (0.00195)	0.12332 (0.08171)
Market price rate	0.02424 (0.04793)	-7.58142** (3.19275)	-0.01833 (0.04593)	-5.18552* (2.05133)
R squared	0.01321	0.33774	0.01103	0.14311
Note. Dep. var.: Direction of closing price and trading volume. Variables represent % changes.				

To incorporate potential coin investors' social belief and response into the model of investment decision, we model the coin market behavior in a Markov process framework that provides a simple and appropriate way of modeling the investment decision conditioning on the current market situation. A Markov chain offers a probabilistic approach in predicting the likelihood of an event based on previous behavior. Methodologically, we apply hidden Markov model (hmm) to our time series data because we could collect sequences of varied lengths to find patterns in past market events combined with social factors.

To determine sequences of reasonable lengths, we analyze when the coin markets are in various regime[3] states, specifically two hidden states of upward and downward in the coin markets. Detecting the regime states are challenging because the number of states is not known as a priori. Moreover, there is not any ground truth on which to train the hidden Markov model. In this context, we conducted two modellings.

1. Model 1, we first examined two hidden states in patterns of high volatility[4] and low volatility of closing price and trading volume because we cannot detect the targeted two hidden states of upward and downward if we cannot first discover high volatility and low volatility in the written algorithm.
2. Model 2, we located four hidden states of high volatility, low volatility, upward, and downward of closing price and trading volume (see bottom graphs in figure 1, 2, 3, and 4).
3. We fitted the hmm with the aimed hidden states of upward and downward for simple reference in the coin markets.

Figure 1 illustrates that the first graph on the top shows what looks like a stationary process punctuated by a few spikes of high volatility. The second one shows a two state hmm to model the process, and the last graph shows a four state hmm on the bottom. The second graph illustrates that for the first 25 days, the coin market was calmer and hence the hidden Markov model has given high posterior probability to Red Line. However, roughly between 25 days and 27 days, the market was volatile, so Black Line had high posterior probability for this period. This has the initial effect of rapidly changing posterior probabilities between the two states. The coin market became calmer approximately between 40 days and 50 days, but additional volatility occurred after 50 days. Thus, the hmm gave high posterior probability to Black Line. Afterward, the markets became calmer again and

[3] In a formal sense, *regime (switching)* is a situation in which *stock market returns* are drawn from two different distributions to describe structural changes in a time series.

[4] High volatility indicates an unstable state, while low volatility indicates a stagnant state

the hmm is consistently giving high probability to Red Line. This is reflected in the increased switching between two hidden states for the hmm.

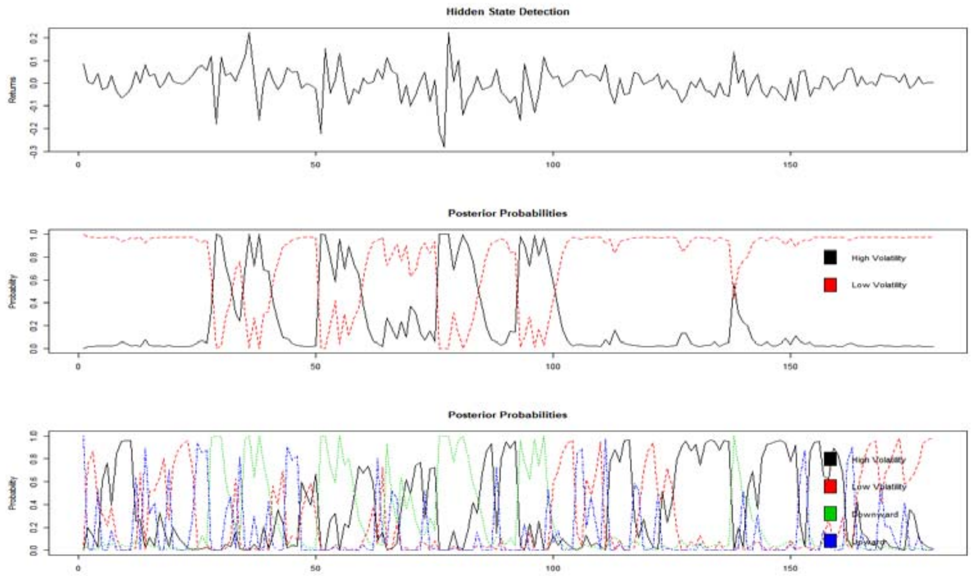


Figure 1. Two and Four Hidden States[5] in the Closing Price of Bithumb

The graph in the bottom demonstrates that the same process was performed for a four state hmm. The model is forced to consider four separate regimes. Thus, it leads to a frequent switching behavior between hidden states. In the first period, Blue Line (upward) dominates the posterior probability, but subsequently the dominance is followed by Green Line (downward). Here, we can observe the model switching between upward and downward with varied sequences. These observations illustrate how to predict the movement direction of closing price and trading volume with what sequences for the coin market behavior.

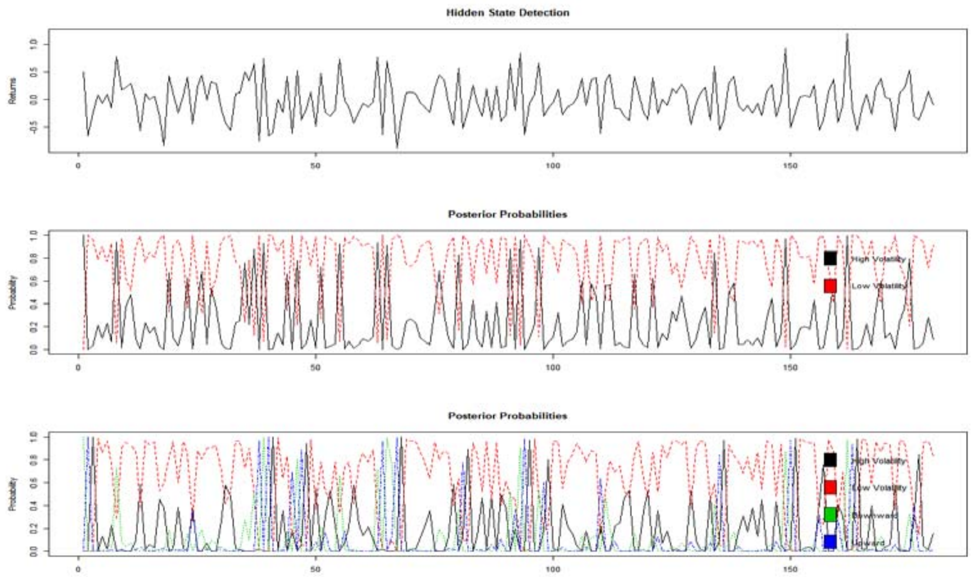


Figure 2. Two and Four Hidden States in the Trading Volume of Bithumb

[5] **Black Line (High Volatility), Red Line (Low Volatility), Green Line (Downward) and Blue Line (Upward)**

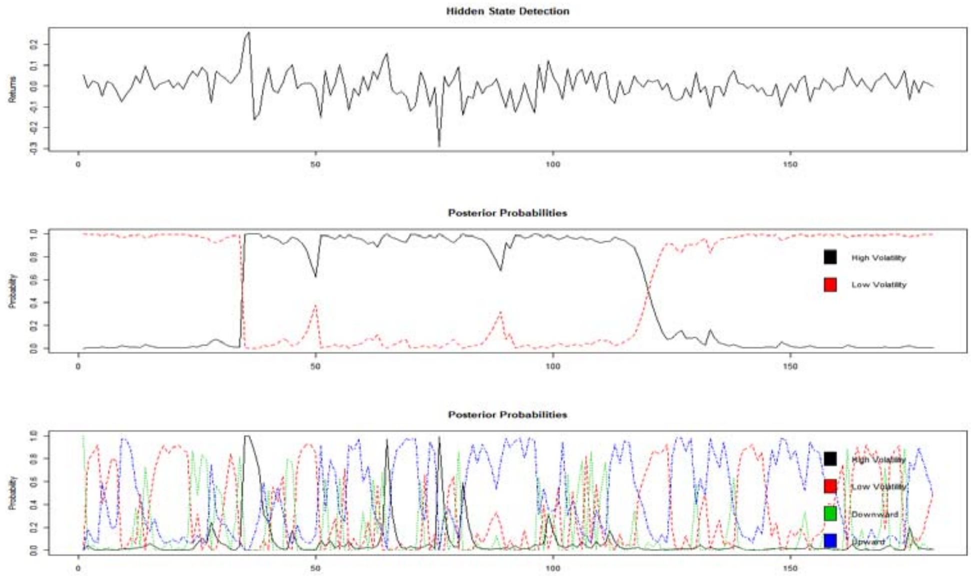


Figure 3. Two and Four Hidden States in the Closing Price of Upbit

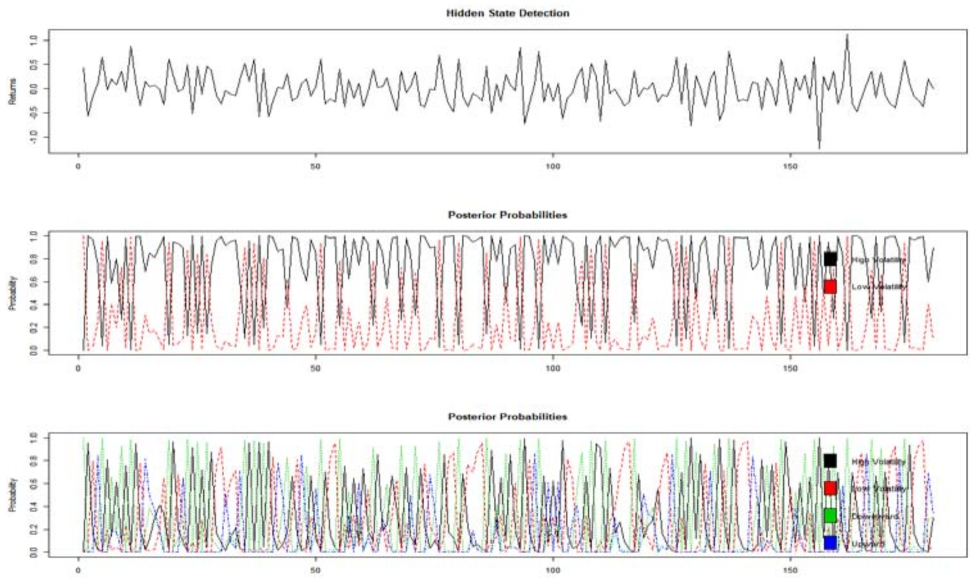


Figure 4. Two and Four Hidden States in the Trading Volume of Upbit

Figure 1, 2, 3, and 4 show similar results, but different sequence patterns. Our data represents one sequence of market events, which leads to the last closing price and trading volume. We split the data set into many sample of sequences to get more sequences and to better understand the market behavior. This way brought about different last closing price and trading volume patterns. We used different periods and controlled varying closing prices and trading volumes. Each sequence as a pattern lead to the movement directions of closing prices ad trading volumes.

To find the appropriate sequence, we used the figure 1, 2, 3, and 4 showing different sequence patterns and adjusted the time length for the sequence. For example, according to a frequent and less frequent switching behavior between hidden states (upward and downward), we set up different time frames e.g., day2, day3, day4 and day5 for a short time pattern or day 2 to day 15 for a long-time pattern. Thus, we had a pattern that matches current coin market conditions using leading closing price and trading volume as indicators for market prediction.

We experimented if the coin market is predicted based on upward or downward market pattern by previous market events and social factors. In figure 1, we modelled returns as a hidden four-state

Markov chain, but we only consider upward and downward states, so that each day's return state is either upward or downward. We algorithmically defined that the hidden state is 1 if return is greater than zero, otherwise 0. We discretized market events and social factors into two value groups (high and low) of equal frequency. Table 4 shows that each variable has only two values of high and low. And then, we simplified all the values combining market events and social factors within a sequence into a single feature.

Table 4. A Single Feature for A Sequence

Variables	Day1	Day 2	Day 3	Day4 (Direction)
Closing price rate	High	High	Low	Trading volume, Closing price
Trading volume rate	Low	High	Low	
Positive sentiment rate	High	Low	High	
Negative sentiment rate	Low	High	Low	
A single feature	HLHL	HHLH	LLHL	Upward or Downward

Now, we create two Markov processes, one for days with upward coin market and another for days with downward coin market. In predicting upward trading volume movement, data includes sequences of observation (closing price rate and trading volume rate). The table 5 shows the transition of the observation (High or Low) and the associated probability.

Table 5. Markov Transition Probability with Trading Volume and Closing Price

Observations	High trading volume Low closing price	Low trading volume High closing price	Low trading volume Low closing price	High trading volume High closing price
High trading volume Low closing price	0.4071923	0.285459	0.2448068	0.0625419
Low trading volume High closing price	0.1373894	0.1225664	0.3097345	**0.4303097**
Low trading volume Low closing price	0.4142186	0.132785	0.1549158	0.2980807
High trading volume High closing price	0.1688552	0.1825412	0.1993712	0.4492325
Note: Observations combine trading volume rate and closing price rate				

Table 5 shows a first order transition matrix from the Markov chain. We simply read the matrix by choosing the previous event on the row and look for the probability on the current event on the column. For example, the highest probability for the current observation after *Low trading volume and High closing price* on the second row is 0.4303097, *High trading volume and High closing price* on the last column.

Table 6. Transition Probability with Four Observations

	LLLH	HHHH	HHLL	LHLL	HHLH	LHHH	HHHL	LHHL	LLHH	HLHL	LLLL	LLHL	HLLH	HLLL	HLHH	LHLH
LLLH	0.1063	0.0540	0.0899	0.0000	0.0958	0.0413	0.0000	0.0000	0.0409	0.0000	0.1176	0.0645	0.0000	0.0672	0.3224	0.0000
HHHH	0.1059	0.3022	0.1046	0.0000	0.0000	0.1440	0.0000	0.0649	0.0839	0.0463	0.0000	0.0532	0.0000	0.0000	0.0467	0.0484
HHLL	0.0582	0.1315	0.0000	0.0759	0.0713	0.0000	0.1177	0.1321	0.0000	0.0000	0.0000	0.1942	0.0000	0.0000	0.2191	0.0000
LHLL	0.0000	0.1908	0.0000	0.0000	0.0318	0.0000	0.4835	0.0000	0.0000	0.0000	0.0000	0.1323	0.0000	0.1399	0.0000	0.0216
HHLH	0.0609	0.2932	0.0000	0.0000	0.0000	0.2072	0.0000	0.0000	0.0000	0.0624	0.1419	0.1742	0.0000	0.0000	0.0602	0.0000
LHHH	0.2083	0.1952	0.0000	0.0000	0.0702	0.0000	0.0000	0.0000	0.0000	0.0000	0.1810	0.0000	0.0000	0.1238	0.2214	0.0000
HHHL	0.2564	0.0000	0.0000	0.0236	0.0524	0.0000	0.1893	0.0797	0.0907	0.0000	0.1400	0.0000	0.0404	0.0000	0.0692	0.0582
LHHL	0.0639	0.0000	0.0000	0.0000	0.0000	0.0000	0.3582	0.4322	0.0000	0.0000	0.1457	0.0000	0.0000	0.0000	0.0000	0.0000
LLHH	0.1391	0.0000	0.0000	0.2880	0.1657	0.1765	0.0000	0.0000	0.2308	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HLHL	0.0000	0.0000	0.0000	0.1246	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.5302	0.0000	0.0000	0.3452
LLLL	0.0000	0.1398	0.0453	0.0543	0.0373	0.0000	0.2875	0.0000	0.1955	0.0000	0.0000	0.0000	0.0987	0.0345	0.1072	0.0000
LLHL	0.0000	0.0000	0.2225	0.0000	0.1597	0.0000	0.1981	0.0000	0.0000	0.0000	0.1326	0.2312	0.0558	0.0000	0.0000	0.0000
HLLH	0.0826	0.2426	0.0000	0.1794	0.0000	0.0000	0.0000	0.3355	0.0000	0.0000	0.1600	0.0000	0.0000	0.0000	0.0000	0.0000
HLLL	0.0239	0.0000	0.2496	0.0000	0.2274	0.0000	0.0000	0.2598	0.0000	0.2103	0.0291	0.0000	0.0000	0.0000	0.0000	0.0000
HLHH	0.1100	0.2114	0.0467	0.0412	0.0000	0.0000	0.0787	0.2790	0.0000	0.0000	0.0922	0.0000	0.0000	0.0000	0.1407	0.0000
LHLH	0.0000	0.0000	0.6338	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3662	0.0000	0.0000	0.0000	0.0000	0.0000
Note: Observations combine positive sentiment rate, negative sentiment rate, trading volume rate and closing price rate																

The following two tables illustrate Markov transition probability for downward trading volume movement.

Table 7. Transition Probability with Trading Volume and Closing Price

Observations	High trading volume Low closing price	Low trading volume High closing price	Low trading volume Low closing price	High trading volume High closing price
High trading volume Low closing price	0.205675	0.441249	0.308132	0.044944
Low trading volume High closing price	0.087189	0.206636	0.39023	0.315945
Low trading volume Low closing price	0.299925	0.286812	0.268877	0.144386
High trading volume High closing price	0.169799	0.221542	0.37698	0.231679

Table 8. Transition Probability with Four Observations

	LLLH	HHHH	HHLL	LHLL	HHLH	LHHH	HHHL	LHHL	LLHH	HLHL	LLLL	LLHL	HLLH	HLLL	HLHH	LHLH
LLLH	0.1711	0.0570	0.1163	0.0000	0.1160	0.0671	0.0000	0.0000	0.0501	0.0000	0.1459	0.0574	0.0000	0.0476	0.1714	0.0000
HHHH	0.1265	0.0887	0.2226	0.0000	0.0000	0.1128	0.0000	0.0639	0.0620	0.0639	0.0000	0.0899	0.0000	0.0000	0.0682	0.1017
HHLL	0.1063	0.0645	0.0000	0.1069	0.1252	0.0000	0.1929	0.0462	0.0000	0.0000	0.0000	0.1935	0.0000	0.0000	0.1645	0.0000
LHLL	0.0000	0.0842	0.0000	0.0000	0.2134	0.0000	0.1740	0.0000	0.0000	0.0000	0.0000	0.1181	0.0000	0.1951	0.0000	0.2152
HHLH	0.1135	0.1010	0.0000	0.0000	0.0000	0.1684	0.0000	0.0000	0.0000	0.1794	0.2701	0.0249	0.0000	0.0000	0.1428	0.0000
LHHH	0.1422	0.1111	0.0000	0.0000	0.2891	0.0000	0.0000	0.0000	0.0000	0.0000	0.1601	0.0000	0.0000	0.1744	0.1231	0.0000
HHHL	0.4376	0.0000	0.0000	0.0728	0.0622	0.0000	0.0874	0.0564	0.0360	0.0000	0.0764	0.0000	0.0521	0.0000	0.0615	0.0575
LHHL	0.1037	0.0000	0.0000	0.0000	0.0000	0.0000	0.3616	0.1700	0.0000	0.0000	0.3647	0.0000	0.0000	0.0000	0.0000	0.0000
LLHH	0.2146	0.0000	0.0000	0.4144	0.1092	0.1663	0.0000	0.0000	0.0955	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HLHL	0.0000	0.0000	0.0000	0.1806	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3326	0.0000	0.0000	0.4868
LLLL	0.0000	0.0834	0.0544	0.0487	0.0653	0.0000	0.2232	0.0000	0.0848	0.0000	0.0971	0.0000	0.1303	0.1890	0.0237	0.0000
LLHL	0.0000	0.0000	0.3763	0.0000	0.1129	0.0000	0.0730	0.0000	0.0000	0.0000	0.1745	0.0331	0.2303	0.0000	0.0000	0.0000
HLLH	0.2803	0.1766	0.0000	0.2584	0.0000	0.0000	0.0000	0.0978	0.0000	0.0000	0.1869	0.0000	0.0000	0.0000	0.0000	0.0000
HLLL	0.2294	0.0000	0.1208	0.0000	0.1427	0.0000	0.0000	0.0963	0.0000	0.1944	0.2163	0.0000	0.0000	0.0000	0.0000	0.0000
HLHH	0.0584	0.1417	0.0880	0.1322	0.0000	0.0000	0.0921	0.1930	0.0000	0.0000	0.2780	0.0000	0.0000	0.0000	0.0165	0.0000
LHLH	0.0000	0.0000	**0.4774**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1794	0.0000	0.0000	0.3432	0.0000	0.0000

Table 8 shows a first order transition matrix from the Markov Chain of four observations. We simply interpret the transition probability from the previous event on the row to the current event on the column. For example, the highest probability for the current observation after *LHLH: low positive sentiment, high negative sentiment, low trading volume and high closing price* on the last row is 0.4774, *HHLL: high positive sentiment, high negative sentiment, low trading volume and low closing price* on the third column.

Results

A common way to describe the performance of a classification model is the confusion matrix[6]. We only consider two classes, upward and downward direction. For two classes, there are additional statistics that may be relevant when one class is interpreted as the event of interest. We split our data set into two subsets, and then used the first four-month data for training[7] dataset and the remaining two-month for test set[8].

Table 9. Confusion Matrix for Two Hidden States Markov Model

Predicted	Observed	
	Downward	Upward
Downward	True Positive (TP)	False Positive (FP)
Upward	False Negative (FN)	True Negative (TN)

[6] This is a simple cross-tabulation of the observed and predicted classes for the data. When the outcome has two classes, diagonal cells denote cases where the classes are correctly predicted while the off-diagonals illustrate the number of errors for each possible case. The simplest metric is the overall accuracy rate. This reflects the agreement between the observed and predicted classes and has the most straightforward interpretation. In Table 9, the top row of the matrix corresponds to samples predicted to be events. Some are predicted correctly, the true positives, or TP while others are inaccurately classified false positives or FP. Similarly, the second row contains the predicted negatives with true negatives, TN and false negatives, FN (Kuhn and Johnson 2013).

[7] A training dataset is a set of examples used to fit the parameters of the model.

[8] A test dataset is a dataset used to provide an unbiased evaluation of a final model fit on the training dataset.

The sensitivity of the model is the rate that the event of interest (Downward) is predicted correctly for all samples having the event,

$$\text{Sensitivity} = \frac{\text{TP (Predicted Downward)}}{\text{TP + FN (Observed Downward)}}$$

The sensitivity is sometimes considered the true positive rate since it measures the accuracy in the event population. Conversely, the specificity is defined as the rate that nonevent (Upward) samples are predicted as nonevents,

$$\text{Specificity} = \frac{\text{TN (Predicted Upward)}}{\text{TN + FP (Observed Upward)}}$$

The most common method for combining sensitivity and specificity into a single value uses the receiver operating characteristic (ROC) curve. The larger the area under the ROC curve (AUC), the more accurate the diagnostic method.

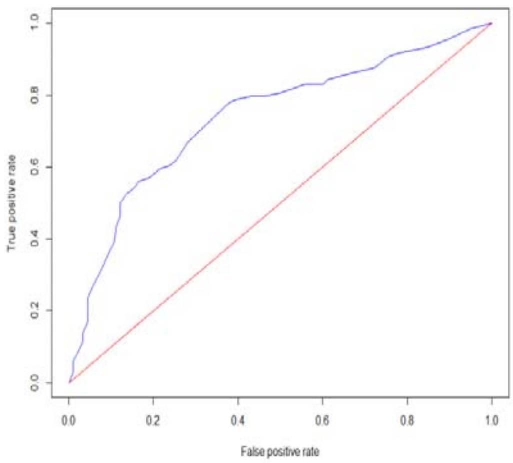
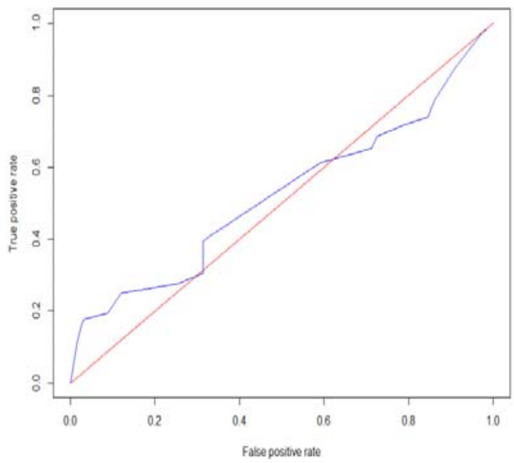
2 observations (trading volume and closing price)	4 observations (volume, price, positive and negative sentiment)
	
Accuracy: 0.6923 Area under the ROC curve (AUC): 0.7388265	Accuracy: 0.5361 Area under the ROC curve (AUC): 0.5219817

Figure 5. ROC Curve for Two Hidden States Markov model

Figure 5 illustrates that previous market events predict the coin market behavior, specifically trading volume direction of Bithumb with a 69% accuracy and a 74% AUC, while combination of market events and social factors less predict the market behavior with a 54% accuracy and a 52% AUC. However, predicting coin market behavior through analyzing the driver of change in the market is interesting.

Table 10. Prediction Results for Two Hidden States Markov Model

Combined Observations for Closing Price Direction (Bithumb)	Accuracy	AUC
Google trend, positive, negative sentiment rate	0.4717	0.5475349
Google trend, positive, negative sentiment, trading volume, price rate	0.4481	0.5576633
Google trend, positive, negative sentiment, trading volume rate	0.3382	0.4962806
Google trend, positive, negative sentiment, price rate	0.5384	0.5760011
Positive, negative sentiment, trading volume, price rate	0.5441	0.5336667
Trading volume, price rate	0.6182	0.6569533
Positive, negative sentiment rate	0.4309	0.5365107
Google trend, positive sentiment rate	0.4702	0.5095169
Google trend, negative sentiment rate	0.5630	0.5332414
Market trading volume, market price rate	0.5018	0.4745814

Table 11. Prediction Results for Two Hidden States Markov Model

Combined Observations for Trading Volume Direction (Bithumb)	Accuracy	AUC
Google trend, positive, negative sentiment rate	0.5507	0.6193125
Google trend, positive, negative sentiment, trading volume, price rate	0.5497	0.5975013
Google trend, positive, negative sentiment, trading volume rate	0.6556	0.7185970
Google trend, positive, negative sentiment, price rate	0.5332	0.5501995
Positive, negative sentiment, trading volume, price rate	0.5361	0.5219817
Trading volume, price rate	0.6923	0.7388265
Positive, negative sentiment rate	0.4898	0.4749873
Google trend, positive sentiment rate	0.4898	0.4749873
Google trend, negative sentiment rate	0.5292	0.5523470
Market trading volume, market price rate	0.4587	0.4141219

Table 12. Prediction Results for Two Hidden States Markov Model

Combined Observations for Closing Price Direction (Upbit)	Accuracy	AUC
Google trend, positive, negative sentiment rate	0.4208	0.4757637
Google trend, positive, negative sentiment, trading volume, price rate	0.4367	0.4864606
Google trend, positive, negative sentiment, trading volume rate	0.3919	0.4882828
Google trend, positive, negative sentiment, price rate	0.5219	0.5432370
Positive, negative sentiment, trading volume, price rate	0.4216	0.5457359
Trading volume, price rate	0.5909	0.5746751
Positive, negative sentiment rate	0.4155	0.4900436
Google trend, positive sentiment rate	0.4550	0.4439350
Google trend, negative sentiment rate	0.4550	0.4439350
Market trading volume, market price rate	0.5710	0.5715191

Table 13. Prediction Results for Two Hidden States Markov Model

Combined Observations for Trading Volume Direction (Upbit)	Accuracy	AUC
Google trend, positive, negative sentiment rate	0.5433	0.6199346
Google trend, positive, negative sentiment, trading volume, price rate	0.4717	0.4819332
Google trend, positive, negative sentiment, trading volume rate	0.4849	0.4942398
Google trend, positive, negative sentiment, price rate	0.4919	0.5226692
Positive, negative sentiment, trading volume, price rate	0.4670	0.4721038
Trading volume, price rate	0.5524	0.5488861
Positive, negative sentiment rate	0.5313	0.5550224
Google trend, positive sentiment rate	0.6047	0.6436695
Google trend, negative sentiment rate	0.5631	0.5424443
Market trading volume, market price rate	0.5009	0.4856802

Table 10, 11, 12 and 13 depict the prediction results of four different experiments in two-coin markets performed in the hidden Markov process framework. First, previous coin market events have a significant role of predicting the coin market behavior in the first three tables e.g., accuracy of coin market events is (0.6182, 0.6569533), (0.6923, 0.7388265), and 0.5909, 0.5746751). However, social factor could be an important leading indicator of future coin market swings in the last table e.g., accuracy of social factors is (0.6047, 0.6436695). Interestingly, stock market events could also have a strong impact on coin market prediction in the second last table, e.g., accuracy of stock market events is (0.5710, 0.5715191).

Discussion and Conclusion

We examine the overall situation of the coin markets and predict the market behavior by combining market events and social factors to incorporate coin investors' social belief and response into the market swings. This is a new study to understand coin market behavior contrast to prior studies (Mai et al. 2018) identifying what determine bitcoin's value and finding the factors (Polasik et al. 2015, Ciaian et al. 2016) that influence its monetary value. We believe that our paper contributes to our understanding of the impact of social metrics on coin investment decisions.

The study results provide evidence of the role of social belief and response in coin investment activity or possibly in decision making in various contexts. We clarified that previous market events are key predictors for the coin market behavior, but not equally in all market conditions. We verified that trading volume rate and closing price rate significantly predict the movement direction of trading volume and closing price through metrics of Accuracy and Area Under the ROC Curve (AUC) in both coin markets of Bithumb and Upbit. However, we identified that Google trend rate and positive sentiment rate perform better the prediction of directing trading volume moves in the coin market of Upbit. In addition, stock market events also fulfill well predicting the movement direction of closing price in Upbit. It is interesting to agonize what is going on in the context that we studied. This study explicitly evaluates the significant role of the market events on the coin market behavior and indicates the importance of social factors even though they are not equally influential in the markets. Thus, we conclude that the 1) market events are overall core indicators and 2) social factors are latently very important but should not be treated in the same way in all coin markets.

In this study, the novelty is that it examines the impact of combined observations of market events and social factors through varied sequences adjusted for each market situation. First, we could analyze coin investors' social belief and response on coin market behaviors in innovative ways. Second, this study specifically extends the cryptocurrency literatures by empirically analyzing coin market movements. Third, many parts shown in our research demonstrated Fintech's supportive role in the coin market and how it will lead to meaningful results.

Although our study examines domains that have not been explored in prior studies, there are places that need to be complemented and developed further.

Our data do not include the market events and social factors associated with other cryptocurrency e.g., Ethereum, Ripple, Monero, etc. Many investors became involved in the coin markets because they expect the value of one or another cryptocurrency to increase. Such collective excitement may be thrown into price bubbles and subsequent market crashes. Each cryptocurrency may lead to a biased or differentiated effect on the analyses and processes in many ways. Thus, the study will be more valuable if further analysis of different coins in the same market is conducted and compared.

Even if the study used data collected over the months, if we could use a lot of data over a longer period, we would be able to uncover previously unexplained phenomena and potential element effects.

We do not explain why the different results are observed in the two-coin markets. One observable point is that in the two markets, the four states (high volatility, low volatility, upward and downward) appear at very different sequences in the return of trading volume and closing price. However, this study does not examine how the latent variables of a coin market (Bithumb or Upbit), a channel (online or offline) or a country (local or foreign) affect the outcomes. If relevant analyses of other specific latent variables are conducted, more valuable and generalizable results might be drawn.

The underlying process that drives the investors' social belief and response and its impact on the coin market behavior is more nuanced. Future research to identify the underlying process will help extend our understanding of the market behavior and social belief of coin investors.

Acknowledgements

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A3A2066740).

Reference

Böhme, R., Christin, N., Edelman, B., and Moore, T. 2015. "Bitcoin: economics, technology, and governance." *Journal of Economic Perspectives* (29: 2), pp. 213–238.

Bouoiyour, J. and Selmi, R. 2015. "What does Bitcoin look like?" *Ann. Econ. Financ* (16:2), pp. 449–492.

Ciaian, P., Rajcaniova, M., and Kancs, D. 2016. "The economics of Bitcoin price formation." *Appl. Econ.* (48:19), pp.1799–1815.

Clemons, E.K., Croson, D.C., and Weber, B.W. 1996. "Reengineering money: The Mondex stored value card and beyond." *International Journal of Electronic Commerce* (1:2), pp. 5–31.

DeLone, W.H., and McLean, E.R. 2004. "Measuring e-commerce success: Applying the DeLone and McLean information systems success model." *International Journal of Electronic Commerce* (9:1), pp. 31–47.

Dewan, R.M., Freimer, M.L., and Zhang, J. 2003. "Management and valuation of advertisement supported web sites." *Journal of Management Information Systems* (19:3), pp.87–98.

European Central Bank. 2012. "Virtual Currency Schemes." Frankfurt am Main.

Evans, D.S. 2014. "Economic Aspects of Bitcoin and Other Decentralized Public-ledger Currency Platforms (article No. 685)." *University of Chicago Coase-Sandor Institute for Law & Economics Research Paper*.

Glaser, F., Zimmermann, K., Haferkorn, M., and Weber, M.C. 2014. "Bitcoin: asset or currency?

Revealing users' hidden intentions." In *Proceedings of the 22nd European Conference on Information Systems*, M. Avital, J. M. Leimeister and U. Schultze (eds.), Tel Aviv: Association for Information Systems.

Garcia, D., Tessone, C.J., Mavrodiev, P., and Perony, N. 2014. "The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy." *J. R. Soc. Interface* (11:99).

Hileman, G. 2016. "State of Bitcoin and Blockchain." New York: CoinDesk, January 28.

Hainmueller, J., and Hazlett, C., 2014. "Kernel regularized least squares: reducing misspecification bias with a flexible interpretable machine learning approach." *Polit. Anal.* (22:2), pp.143-168.

Jin, S.-A.A., and Bolebruch, J. 2009. "Virtual commerce (v-commerce) in Second Life: The roles of physical presence and brand-self connection." *Journal of Virtual Worlds Research* (2:4).

Kristoufek, L. 2013. "Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era." *Scientific Reports* (3), pp. 1–7.

Kristoufek, L. 2015. "What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis." *PloS ONE* (10:4).

Kuhn, M., and Johnson, K., 2013. "Applied predictive modeling." Springer New York Heidelberg Dordrecht London

Jopson, B. 2016. "Regulators say bitcoin poses financial stability risks." *Financial Times*, June 21.

Li, X., and Wang, C.A. 2017. "The technology and economic determinants of cryptocurrency exchange rates: the case of Bitcoin," *Decision Support Systems* (95), pp. 49-60.

Mai, F., Shan, Z., Bai, Q., Wang, X., and Roger H.L. Chiang. 2018. "How does social media impact bitcoin value? A test of the silent majority hypothesis." *Journal of management information systems* (35:1), pp.19-52.

Polasik, M. Piotrowska, A.I. Wisniewski, T.P. Kotkowski, Radoslav. Lightfoot, G. 2016. "Price fluctuation and the use of bitcoin: an empirical inquiry." *International journal of electronic commerce* (20: 1), pp. 9-49.

Ren, S., and Culpan, T. 2017. "Ethereum's wild ride needs to slow." *Bloomberg Businessweek*, July 13.

Tapscott, D., and Tapscott, A. 2016. "Blockchain Revolution: How the Technology Behind Bitcoin is Changing Money," *Business and the World*, 2016 (Portfolio).

Yermack, D. 2013. "Is bitcoin a real currency? An economic appraisal." *NBER Working Paper*, p. 19747.

Zähres, M. 2012. "E-money. Niche market that might be expanding. Deutsche Bank Research." *Banking & Technology Snapshot*, May 11.