

The Effect of Tweets Made by Cryptocurrency Opinion Leaders on Bitcoin Prices

Shakirullah Hamza*

Shanghai University, 99 Shangda Road, BaoShan district, Shanghai, China

DOI: [10.36348/sjef.2020.v04i12.005](https://doi.org/10.36348/sjef.2020.v04i12.005)

| Received: 27.11.2020 | Accepted: 09.12.2020 | Published: 13.12.2020

*Corresponding author: Shakirullah Hamza

Abstract

Rapid technological advancements in the last few decades have given rise to various new products and fields, such as cryptocurrencies, social media and sentiment analysis. The massive surge in internet usage has caused organizations and investors to increasingly base their decisions on content placed on social media platforms which are flooded with data from its users. One of those platforms is twitter, a micro-blogging platform which allows people to share their opinions in a limited number of characters. Certain users on these platforms have the ability to influence other users' decision making, including investing. Although stock market prediction through sentiment analysis has been researched often, the amount of research done on prediction Bitcoin prices is relatively low. Furthermore, there is a gap in existing research in which samples are not limited to users that have more knowledge than individual investors. This paper uses sentiment analysis on tweets made by cryptocurrency influencers to see whether they can be used to predict Bitcoin price fluctuations. This paper, additionally researches differences in predictive capabilities between regions and differences in predictive capabilities between a bear and a bull market. The results indicate that tweets made by cryptocurrency influencers contain statistically significant information about the future value of bitcoin. Thus, analyzing those tweets to exploit profitable opportunities can be worthwhile.

Keywords: Bitcoin, Cryptocurrencies, Sentiment analysis, Twitter, Microblogging, Times series analysis.

Copyright © 2020 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC 4.0) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

INTRODUCTION

The recent technological revolution has led to numerous advances in various fields, among which are advances in the field of finance, communication technologies and advances in data analysis, specifically sentiment analysis. In the field of finance, this revolution has introduced the concept of cryptocurrencies, digital currencies in which encryption techniques are used to regulate the generation of units of currency and verify the transfer of funds, operating independently of a central bank [1]. Cryptocurrencies are a globally spreading phenomenon which are gaining an increasing amount of attention from media, governments and financial institutions.

The first cryptocurrency to make its appearance was Bitcoin which was introduced in 2008 by an unknown group or individual going by the name of Satoshi Nakamoto [2]. Since its inception Bitcoin has been presented as a decentralized, borderless method of transferring funds and over the years its adoption has brought this statement closer to reality. Currently there are over 400 [3] exchanges on which Bitcoin can be

traded, and these are placed all over the world. In the short time span since its inception in 2008, Bitcoin has reached various corners of the world. Furthermore, exchanges such as the NASDAQ have plans on incorporating Bitcoin, which would further legitimize and increase Bitcoin's adoption rate.

Additionally, its quick adoption rate has attracted many speculators and its price has experienced extreme amounts of volatility, seeing an increase from \$0.3 in January 2011 to \$1,300 in November 2013, experiencing an all-time-high price of \$20,000 and dropping back again to \$3,200 in December 2018 (See Figure-A3). These fluctuations in price not only caught the attention of regulators and institutions, but also those of active traders, retail traders and investors seeking profitable opportunities.

Some of these traders and investors have created a following on social media platforms. Because of the technological revolution and the increasing available of the internet, social media platforms have experienced a surge in usage, leading to organizations and individuals using the content on these platforms to

make decisions [4]. One of the new forms of communication is microblogging in which users can "describe current status in short posts distributed by instant messages, mobile phones, email or the Web" [5]. One of, if not the most, used platforms which provides microblogging is Twitter, a social networking service where people can communicate through short messages, called tweets, which are limited to 280 characters. These tweets are generally used to make announcements and post opinions.

OBJECTIVES OF THE PAPER

Firstly, just as being able to value, predict stock prices and knowing which factors affect them, predicting cryptocurrency prices could lead to profits for individuals as well as companies and perhaps affect monetary policies.

Secondly, because of cryptocurrencies and sentiment analysis being relatively new technology, the literature on them is limited. Thus, this paper's second purpose is to extend existing literature on cryptocurrencies and sentiment analysis. Furthermore, current research has mainly focused on the overall sentiment of Twitter on stock markets, taking into account information spread by bots as well as internet trolls and additionally ignoring that most of the Twitter users place uninformed opinions and that opinion leaders have a significant influence on the investment decisions of others. As mentioned before, analyzing the tweets of opinion leaders might prove to be a better source of predicting stock prices as they are able to analyze stock markets more deeply than other Web users. This paper will make an attempt to fill this gap and thus its research question is whether Bitcoin prices can be predicted using the tweets of Twitter opinion leaders. In addition, even though cryptocurrencies and twitter are accessible worldwide, they are not to the same extent. Twitter users per region differ and countries in the Asian region have experienced premiums on Bitcoin prices, thus it would be interesting to study which differences there are on the effects of twitter opinion leaders between regions. Furthermore, as bull and bear markets have a different effect on investors, it is worth looking into whether or not there is a difference in the effect of tweets in both markets.

To achieve this goal, the paper uses a quantitative research method and conducts a time series analysis. Specifically, for this research the Vector autoregression model is used, whose results are afterwards summarized using Impulse responses and Grangers causality.

The results indicate that the sentiments of tweets impact return and volatility and they that there is a Granger-cause relationship. And although sentiment does not Granger-cause volume, an impulse to sentiment does have an impact on volume.

LITERATURE REVIEW

Bitcoin and Cryptocurrencies

Cryptocurrency is a decentralized virtual currency, based on blockchain technology, which uses cryptography for security. The most well-known cryptocurrency is Bitcoin, which was first introduced in 2009 by Satoshi Nakamoto [2].

Unlike fiat, currencies issued by traditional central banks, Bitcoin's supply is limited as there can only be 21 million Bitcoins in existence. The value of Bitcoin is not backed by gold or other commodities, but it's based solely on a value that people assign to it.

A number of studies that have been done on estimating Bitcoins fair value have resulted in conclusions of Bitcoin being mainly priced through speculation. In their works, Fantazzini et al., [6] reviewed models for deriving fundamental value through methods of market sizing and marginal costs of production based on electricity consumption and reviewed proposed econometric models which used cross-sectional, univariate and multivariate time series models. They conclude that speculation is the main driver of Bitcoin prices, followed by supply and demand related variables. They lastly conclude that global macro financial variables play no role. These findings coincide with the results of Liu and Tsyvinski [7]. They call into question explanations such as mining costs and price-to-"dividend" ratio being useful for predicting cryptocurrency returns. They further conclude that the returns of cryptocurrencies can be predicted through time-series momentum and investors' attention which is measured through google searches and tweets on twitter. Cheah and Fry [8] too conclude that Bitcoin prices contain a considerable speculative component and that its fundamental value is zero.

Microblogging and Twitter

One of the reasons of Bitcoins increasing popularity is social media. Over the years social media's integration into daily lives has increased at an astonishing rate. Currently there are 3.484 billion active social media users which exceeds the predicted number of 2.77 billion in Figure-A4 [9].

Currently 69% of us adults and 88% of Americans aged 18-29 use social media, spending two and a half hours daily on social media platforms [10]. One form of communication is called microblogging through which users can, in short messages, share their opinions. One widely used platform that has been used in various studies to mine data from is Twitter. Twitter has 321 million monthly active users which is visited daily by approximately 46% of its users [10]. Other than the number of users, twitter contains users that vary from celebrities, company representatives and as twitter is used worldwide, it is possible to get data in multiple languages [11]. These factors make twitter a rich source of data for sentiment analysis

Sentiment Analysis in Forecasting

This integration into aspects of daily lives and the increasing use of social media has led to a huge volume of recorded opinions. This growth of social media coincides with the rapid growth of a field named sentiment analysis. Liu [4] defines sentiment analysis as "the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes".

The growth in the field of sentiment analysis has caused sentiment analysis to be applied to forecasting.

A variety of studies have been performed on using sentiment analysis on micro blogs to predict movements in the stock market. Bollen et al., [12] for instance showed in their study that when days contained a lot of tweets with, according to Google-Profile of Mood States (GPOMS), a "calm" sentiment, the DJIA rose over the new few days. The study of Oh and Sheng [13] shows that micro blog sentiments contain information that can be valuable to investment decision making and that people do in fact use social media to influence their investment decisions. Another study conducted by [14] implies that the sentiment analysis of public mood can be used for forecasting individual stock prices. The event study conducted by Ranco et al., [15] shows that although the period of impact of Twitter sentiment on price returns is lower for only non-earnings announcements events, the impact is still statistically significant. The results of all these studies imply that sentiment analysis can indeed be used for predicting stock market movement.

Just like on stocks, studies on sentiment affecting cryptocurrency have also been carried out, however in lower volume, which have resulted in mixed conclusions. [16] conclude that social media can be used use to predict future Bitcoin returns as a shock of positive postings indicate positive returns the next day and a negative shock indicates a negative return the next day. Additionally [17] and [18] were able forecast prices with an over 75% accuracy. Whereas [19] neglected the sentiment analysis because of the lack of relationship between sentiment and price. They concluded that even when prices decreased the sentiment was still high. The study conducted by [20] shows a moderate correlation of Twitter sentiment with close price and volume. Furthermore, "Granger causality analysis showed no statistical significance for Twitter sentiment as a predictor of close price, intraday spread or intraday return" Kaminski [20]. By conducting a time series analysis [21] were able to provide evidence of sentiment analysis being able to predict movements in the short-term.

Opinion Leaders

As mentioned before, when making decisions people take into account the opinions of others. Not everyone, however, has the same influencing power. The opinion of some people is held in higher regard than those of others, and the case of stocks is not an exception. These people are referred to as opinion leaders. As cited by Duan et al., [22] "Especially, in network stock forums, opinion leaders are usually able to analyze stock markets more deeply than other Web users, and thus their opinion about the movement of stock price might be valuable for many investors".

A recent example of an algorithm using twitter sentiment to make trades based on the tweets of an opinion leader is called TRUMP2CASH [23]. This algorithm, which trades based on trumps tweet, has according to its historical benchmark, an annualized return of 59% since its inception. These results indicate that the effect of the tweets of influential people are worth looking into as a predictive element. This results in the following hypotheses are created:

H1a: The sentiment of cryptocurrency opinion leaders on twitter can be used to predict Bitcoin prices

H1b: The sentiment of cryptocurrency opinion leaders on twitter can be used to predict Bitcoin volatility

H1c: The sentiment of cryptocurrency opinion leaders on twitter can be used to predict volume

Opinion leaders play a crucial role in online communities as they have enough influence to guide the direction of public opinion. And as stock prices as well as cryptocurrency prices are impacted by public opinion, influential leaders have the ability to move the market.

Although in general, the research on sentiment analysis forecasting cryptocurrency prices bears mixed results, limiting the sample to opinion leaders seems to increase predictive accuracy.

As mentioned by Sul et al., [24], twitter followers with many followers, such as Jim Cramer who has 650.000 followers, have a market impact similar to traditional media, meaning that their sentiment is rapidly incorporated into stock prices.

Another example of is a study on the tweets of Trump that show that the sentiment of a tweet made by Donald Trump can significantly affect financial markets and that the impact of a positive tweet is stronger than the impact of a negative tweet. The study further concludes that its findings are inconsistent with the semi-strong market hypothesis as the effects of the tweets last up to three days [25].

When analyzing tweets Mai et al., [16] found that the impact of bullish and bearish tweets made by all the users had no significant impact on Bitcoin returns in

the next hour. But by limiting sample to those with a lot of followers the relationship became more significant.

Furthermore, as Bitcoin is still arguably in the early stages of its development, most of its users are to a certain extent tech-savvy and highly likely to be using the social media.

Geographical Region

The second hypothesis that this paper will test is:

H2: The predictive ability of the sentiment of a tweet is more accurate during NYSE trading hours

Although Bitcoin as well as Twitter are globally available, most of twitter's users, including the opinion leaders, reside in the United States and thus it is expected that traders in the U.S. will be the first to react those tweets. Moreover, according to the Efficient Market Hypothesis, the sentiment of the tweets should be immediately shown in the price of Bitcoin.

Additionally, the research conducted by Makarov and Schoar [26] on arbitrage shows that arbitrage spreads within regions are small whereas arbitrage spreads between regions can be large. Over most of 2017 Korean and Japanese exchanges had huge premiums over US exchanges which insisted for long periods of time. The premium on a Korean exchange peaked at 1.6 times that of the US and the premium of a Japanese exchange peaked at 1.2 times that of an US exchange.

Difference in bearish/bullish market

The opinions of investors in bullish and bearish markets have a different effect which leads us to the third hypothesis: H3: The effect of an impulse in a bear market is higher

A bullish market is a market in which prices are rising or expected to rise. Whereas the bear market is a market in which prices are dropping or are expected to drop. Bullish sentiment is often related to overconfidence and wishful thinking which could lead to speculative bubbles, such as the dotcom bubble and arguably Bitcoin.

Oh and Sheng [13] argued in their paper that virtual investment communities have the characteristics of containing traders "who each hold a small bit of information" and "have to rely on inferences from observed behavior in order to estimate asset values" and are thus highly susceptible to wishful thinking. The same applies to Bitcoin markets. Although Bitcoin's adoption has increased rapidly over the year, it is still arguably still in its infancy state, is still a highly speculative asset [27] and does not have a method on how to derive its value other than see what value others assign to it.

Although, people are more impacted by bad news rather than good news [28] and as cited in Oh and

Sheng [13] "according to prospect theory Kahneman & Tversky, 1979, losses weigh more than gains", people in the Bitcoin market have experienced massive gains as well as losses of up to 90% due to its volatility. Adding to that the fact that many of its supporters distrust central authorities and could be investing Bitcoin irrespective of its financial gains, the sentiment could be positive even in bear markets, as was also noted by Abraham et al. [19].

METHODOLOGY

Data gathering and Processing

To successfully perform this research three sets of data were necessary. The first set of data that was needed was a list of opinion leaders on twitter which were active in the cryptocurrency space. The second set of data that was needed were the tweets of said opinion leaders. And the last set of data that was needed was the historical Bitcoin prices data.

Twitter Data

Because of what the users post, real time messages about their opinions on a variety of topics, there is large amounts of data that can be gathered from microblogging websites, making them very attractive for sentiment analysis [11]. Furthermore, according to Kouloumpis et al., [29] showed that features of microblogging, such as presence intensifiers, emoticons and abbreviations proved to be most useful for sentiment analysis.

Twitter was chosen to get data from, because it's the biggest microblogging service in the world, with more than 126 million [30] daily users and over 500 million [31] tweets per day and 200 billion tweets per year.

To gather enough data on tweets about Bitcoin, the choice was made to include influential, but also active, participants in the cryptocurrency space. For this reason, highly influential figures (e.g. Donald Trump) were not taken into account. With the help of google, a list of the top 100 [32] cryptocurrency twitter opinion leaders was acquired. Of the top 100, the tweets of the top 50 people were mined. The list contains twitter users that are highly affiliated with the cryptocurrency scene, such as CEO's, venture capitalists and active traders. Gathering the tweets of influential people helps with 2 things. Firstly, by choosing a specific list of people we avoid tweets made by bots, which are estimated to generate approximately 24% [33] of all tweets on twitter. Especially in the cryptocurrency space bots are very prominent and known to be used by scammers. Secondly by taking a list of active opinion leaders on twitter, we can assume that they have at least some knowledge on what they are talking about and thus their tweets might provide some information to be incorporated in the price.

Because of the limitations of the Twitter API which allows access to historical twitter data for up to 7

days, a third-party python project, going by the name of Get Old Tweets [34], was used to mine the tweets. Personal changes to the project were made to ensure the gathering of Bitcoin related tweets by only including tweets that contained the following keywords: Bitcoin, btc, #btc, \$btc, \$Bitcoin, #Bitcoin, Bitcoin., btc., #btc., \$btc., \$Bitcoin., #Bitcoin, and to account for case sensitivity, the tweets were all turned to lower case. From each individual tweets data such as date, time, username, tweet content, retweets, favorites was gathered.

Variables

In the time-series analysis the following variables are used and the rest is omitted as those were not relevant to the hypotheses.

- **Return:** The difference between the Bitcoin Open price at the start of the interval and the Close price at the end of the interval.
- **Volatility:** Absolute value of the Return.
- **Sentiment value:** The sentiment value is the average sentiment value of the tweets acquired by VADER within the interval.
- **Volume:** The volume, in BTC, within an interval

Sentiment Value

To get the sentiment value for each tweet, a sentiment analytic software, called VADER (for Valence Aware Dictionary for sEntiment Reasoning) [35], was used. VADER uses a list of "lexical features, specifically aimed at microblogging, combined with five general rules that embody grammatical and syntactical conventions for expressing and emphasizing sentiment intensity". It outperforms eleven other highly state-of-practice benchmarks (e.g. LIWC, ANEW and the General Inquirer [35]). Vader not only classifies each tweet as positive, negative or neutral but also assigns a value to the sentiment between -1 and 1 with -1 being most extreme negative and 1 being

Bitcoin Historical Data

The reason Bitcoin was chosen over any other cryptocurrency is, because it's the cryptocurrency that serves as a gateway into other cryptocurrencies. Not only is it the first decentralized cryptocurrency, it is also the most well-known, most talked about and most liquid.

In this research the historical 1Min¹ Bitcoin Data is acquired. The data is drawn from the Gemini Exchange. Although Gemini is not the most used exchange, this was the only reliable source to get Bitcoin/USD data from on a 1-minute basis.

Resampling

Data was gathered in the range of January 1st 2017 to the 23rd of April 2019. This was done because

firstly, in the years between 2010 and 2017 the number of tweets were very low.

Secondly taking the range of data from 2017 onwards has the advantage of including both a bull and a bear market. The bull market started around early 2017 and lasted until mid-December 2018 in which the price saw an increase from around \$1,000 to \$20,000 and saw a drop back to around \$3,000 mid-December 2018.

After matching the time zones of the data, to test whether a change in data intervals could result in better forecasting accuracy, the data was resampled into data with intervals of 15 minutes, 30 minutes and 60 minutes. In those intervals, the return is the difference between the open price and the close price for the corresponding interval and the sentiment score was the average of all the sentiments of the tweets in the interval.

Because the variables are of different scale and to avoid giving variables with larger ranges more importance in the analysis, the data was standardized by subtracting the mean and dividing by the standard deviation.

The data was then limited to active trading hours of three major stock exchanges across the world, because as mentioned by Baur et al., [36], trading activity lowered in the weekend and during local evening hours. The exchanges chosen were the New York Stock Exchange, the London Stock Exchange and the Tokyo Stock Exchange.

Analysis

To measure the forecasting ability of the variables, the Vector Auto regression model is used. According to Zivot and Wang [37] the VAR model is one of the most successful and easiest to use model for multivariate time series analysis. Furthermore, the var model can be used in structural analysis in which "certain assumptions about the causal structure of the data under investigation are imposed, and the resulting causal impacts of unexpected shocks or innovations to specified variables on the variables in the model are summarized" [37]. These causal impacts can be summarized through impulse response functions and Granger causality. Impulse responses are estimated responses of variables to impulses on other variables in the system. If a variable reacts to an impulse in another variable, it could be said that the former is caused by the latter variable. Granger causality tests whether one time series is useful in forecasting another time series.

The VAR analysis was conducted using python's StatsModel [38] package and estimates the relationships between the time series and their lagged values with the methods of Lutkepohl [39]. The lag order, automatically chosen by the function, equals to 1.

¹Data retrieved from:
<https://www.cryptodatadownload.com/data/>

An additionally augmented Dick-Fuller test is conducted to test for the stationarity of the data.

Its results are then summarized with StatsModels' orthogonalized Impulse Response Function, with the orthogonalization being done using the Cholesky decomposition of the estimated error covariance matrix. Because of this, the variable ordering might change the interpretation. Thus, the variables that are believed to cause the others, are placed first. In this order is as follows: sentiment, return, volatility, volume. Mainly the variable sentiment is believed to cause other variables because, if the sentiment is high or low it will cause people to trade and thus affect returns, volatility and trading volume.

RESULTS

Effect on return, volatility and volume

To test for differences in using different time intervals, this paper uses the data corresponding to the region's trading hours which has the most trading volume. Because most Twitter users reside in the US the data of corresponding to the hours of the NYSE is used to proceed.

The results of the augmented dick-fuller test conducted on the 1-minute interval data indicate that all the variables are stationary at $\alpha = 0.05$, with the resulting p-value being 0.

Comparing the IRF of the 15 min interval data with those of the 30 minute and 60-minute interval (See Figure-1), especially the confidence intervals of the 30 minute and 60-minute intervals are inaccurate as the confidence intervals are well below zero. The confidence intervals get really tight in the second period on all the time intervals.

Although the volatility seems to be affected in a similar way to an impulse to sentiment, the effect on volatility on the 15min interval data bear more accurate results. Furthermore, in the first period of the 15Min interval the impact instantly declines, whereas on the 30and 60Min intervals, the impact increases in the first period and ends up lowering the periods after.

The results of the Granger causality test in Table-1, rejects the Null hypothesis, meaning that Sentiment provides statistically significant information about the future values of return. Therefore, hypothesis H1a is accepted.

Table-1

| H_0: Granger causality test on whether sentiment causes return | | | | |
|--|--------|----------------|---------|---------------|
| alpha | t-stat | critical value | p-value | Reject/Accept |
| 0.05 | 12.126 | 3.842 | 0.000 | Reject |

In general, in the 15-minute data interval IRF of all regions (See Appendix A, B and C), tighter confidence intervals are observed and thus yield more accurate results.

As can be seen in Figure-1, an impulse to sentiment has a positive effect on volatility with the effect declining over the first four periods. The results are significant as the confidence intervals don't dip below zero.

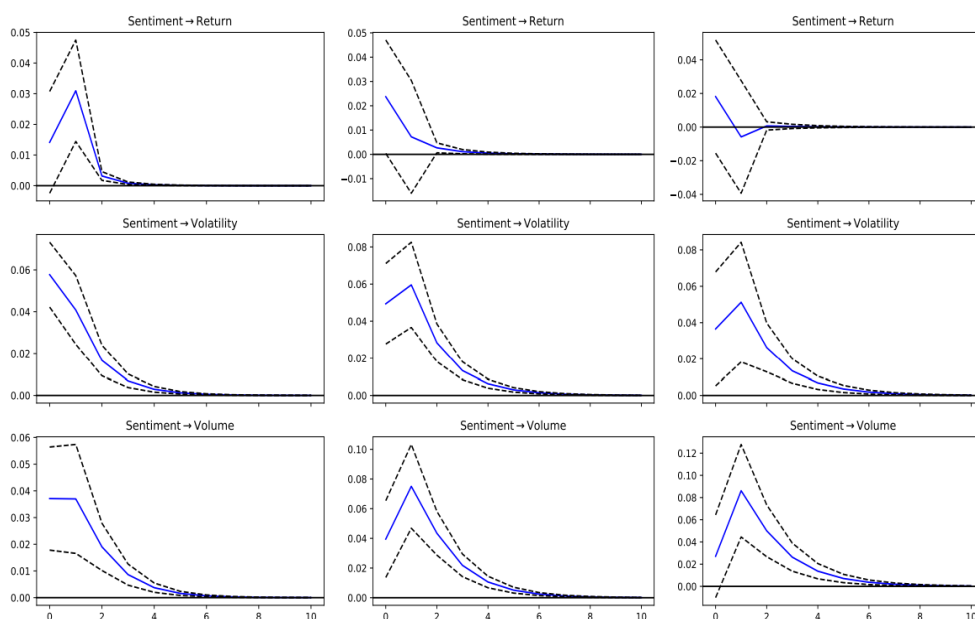


Fig-1: Orthogonalized impulse response functions of the Sentiment, with time intervals 15Min/30Min/60Min respectively, based on data taken corresponding to the trading hours of the NYSE

Furthermore, the results of the Granger-causality test, shown in Table-2, conclude that sentiment Granger-causes Volatility. Because of the

significant impulse response and results of the Granger-causality, hypothesis H1b is accepted.

Table-2

| H_0: Sentiment does not Granger-cause Volatility | | | | |
|--|--------|----------------|---------|---------------|
| alpha | t-stat | critical value | p-value | Reject/Accept |
| 0.05 | 6.563 | 3.842 | 0.010 | Reject |

An impulse to Sentiment leads to an increase in volume (see Figure-1) which is significant as the confidence intervals, don't go below zero. Table-3,

however shows that sentiment does not Granger-cause volume as H0 is rejected.

Table-3

| H_0: Sentiment does not Granger-cause Volatility | | | | |
|--|--------|----------------|---------|---------------|
| alpha | t-stat | critical value | p-value | Reject/Accept |
| 0.05 | 2.056 | 3.842 | 0.152 | Accept |

Furthermore, doing a Granger causality test the other way around shows that volume does Granger-cause Sentiment (Table-4). Therefore, H1c is rejected.

Table-4

| H_0: Sentiment does not Granger-cause Volatility | | | | |
|--|--------|----------------|---------|---------------|
| alpha | t-stat | critical value | p-value | Reject/Accept |
| 0.05 | 8.230 | 3.842 | 0.004 | Reject |

Different Regions

Because the 15Min interval data shows more accurate intervals, the test for regions will use the 15Min data. The impulse response function, in Figure-2, show that regarding the effect on return, in the first period cannot be statistically distinguished from zero. Regarding the volatility, Figure-2 shows very similar results in the effects on the trading hours of the LDSE

and the NYSE, whereas the TKSE's confidence interval goes below zero and thus the impact is statistically insignificant in the first period. Because of the high confidence intervals of the effect on volatility on the TKSE exchange, H2 is accepted with regard to TKSE. But the results between the NYSE and LDSE are not very different so H2 is rejected with regard to the NYSE.

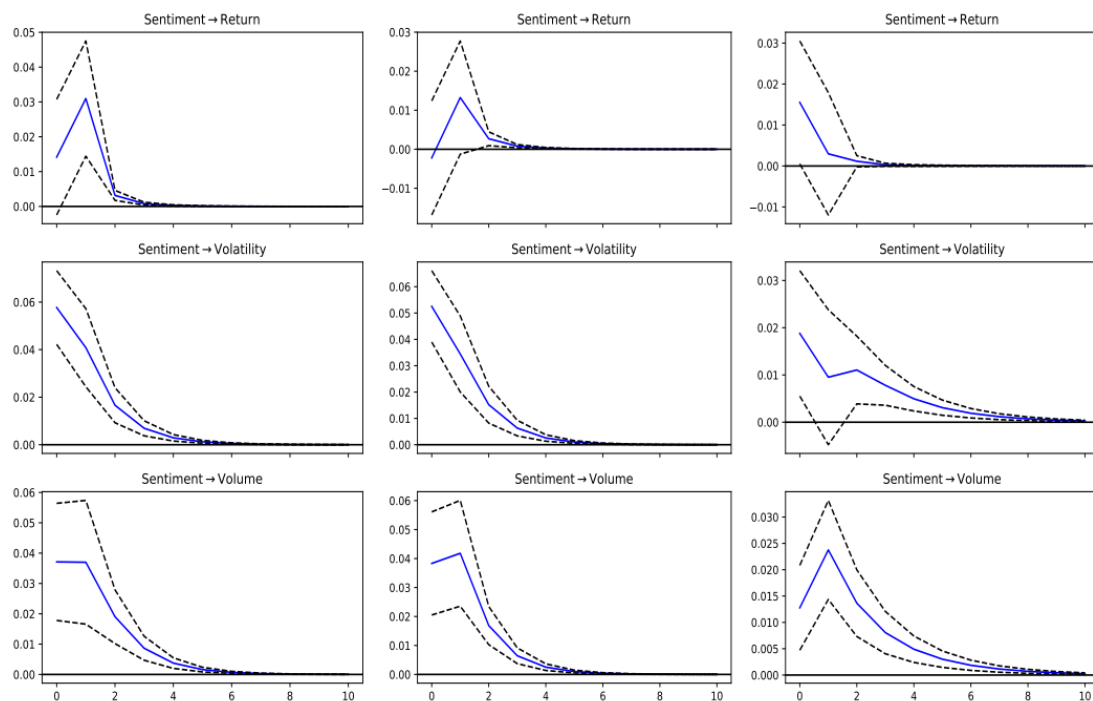


Fig-2: Orthogonalized impulse response functions with trading hours of NYSE, LDSE, TKSE respectively, with 15-minute interval data being used

Bull/Bear Market

The impulse response function graphs in Figure-3 show that the effect on volatility and return in both the bull and bear market is significant, but in the bull market there is a higher impact on both variables. For return the expected changes are 0.02 and 0.00 in the first period of the bull and bear market respectively. For

volatility the expected changes in the first period is 0.08 and 0.03 in a bull and bear market respectively. The same holds for volume, the expected change of 0.03 in a bull market is higher than that of 0.01 in a bear market. Because of the higher expected changes show in figure 3, hypothesis H3 is accepted.

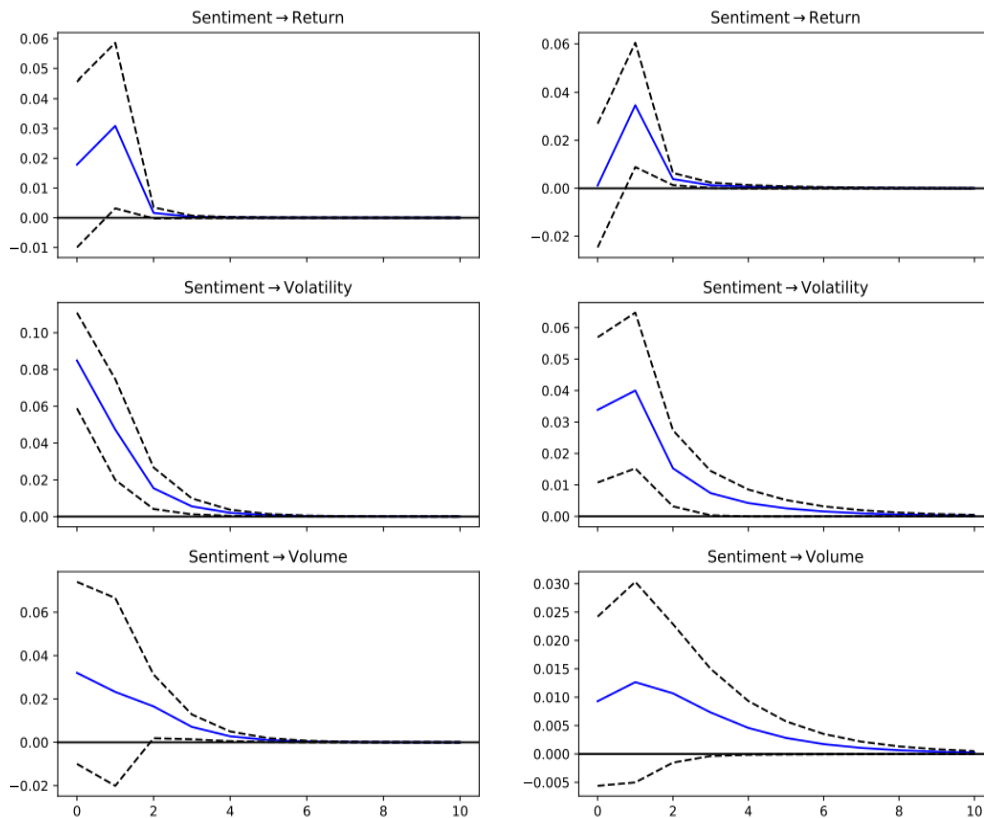


Fig-3: Orthogonalized impulse response functions of the sentiment score on return, in a bull and bear market respectively, with 15-minute interval data being used with trading hours corresponding to those of the NYSE

DISCUSSION AND CONCLUSION

This paper researched the relationship between the sentiment of tweets made by cryptocurrency opinion leaders and Bitcoin returns, volatility and volume.

According to the results of the IRF and the Granger-causality tests, sentiment has an effect on returns of Bitcoin prices, its volatility and volume and thus that tweets contain useful information concerning the future value of Bitcoin. Therefore, tweets made by cryptocurrency opinion leaders contain elements which can be used to predict Bitcoin prices. These results coincide with previous research of sentiment analysis being performed on tweets about stocks [14, 12, 15] as well as on cryptocurrencies [16, 17]. A Granger causality analysis conducted by Kaminski [20] contradicts with the results of this research in that it concludes that sentiment does not Granger-cause Bitcoin price. The results could be different because of the simplistic sentiment analysis of tweets performed by Kaminski [20]. Furthermore, the tweets were gathered

from random users instead of those with some expertise in the market.

There is a clear relationship between sentiment and volatility. The results of both the IRF and the Granger causality test show that there is a relationship between sentiment and volatility.

Although an impulse to sentiment seems to have an impact on volume, the Granger causality results in non-rejection of the null hypothesis. This could be interpreted as increased trading volume causing price changes and as a result lead opinion leaders to publish their opinions.

The results of the differences between regions was expected as there is some overlap between the trading hours of the NYSE and the LDSE, but there is no overlap with the trading hours of the TKSE. Furthermore, twitter is more used in Europe and the U.S., so there would be very few trades made based on the opinion of, mainly, North American twitter users by

the Japanese twitter users. Additionally, the Bitcoin prices data comes from a US based exchange which might have restrictions for international users. And as mentioned by Makarov and Schoar [26] there exist arbitrage opportunities between regions.

Lastly the higher accuracy in the bull market could be caused because by people being optimistic regardless of the price action. These results coincide with the results of Abraham et al., [19] whose results showed that sentiment analysis for cryptocurrencies is less effective in a bear market. Additionally, as Bitcoin has experienced bear markets, that caused the price to go down by 90%, in the past and still bounced back, investors might be willing to hold their Bitcoins regardless of the losses they incur at the time and staying optimistic.

In the bear market, the overall trading volume activity on twitter decreases. This lower activity on twitter could be a possible explanation of the effects of impulses lasting longer, as lower activity results in lower amount of people reacting to the tweets.

In conclusion, the tweets made by cryptocurrency influencers contain statistically significant information about the future value of Bitcoin. This information can then be used to create profitable trading strategies. Lastly the results of this thesis contribute to existing literature on predicting Bitcoin prices and fills the gap of using sentiment analysis on tweets made by opinion leaders as well as

strengthening the use case for using sentiment analysis in forecasting.

LIMITATIONS AND FUTURE RESEARCH

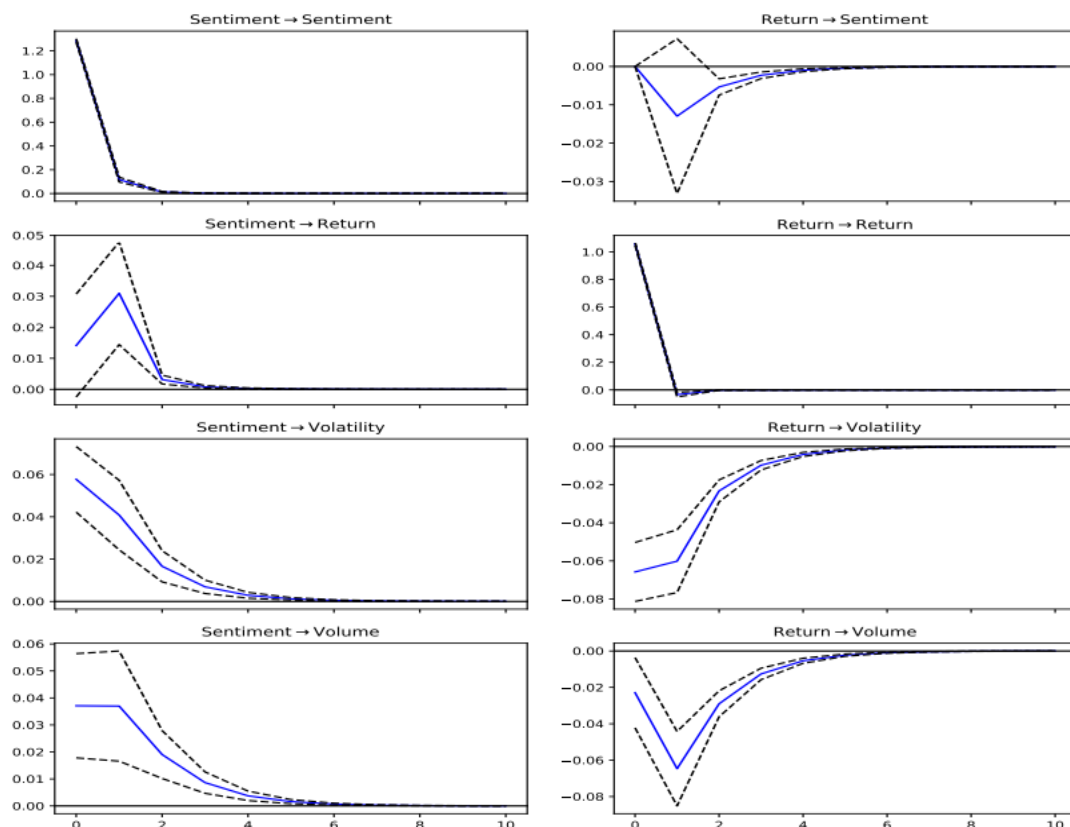
Because of the relatively low age of Bitcoin this study was performed under some limitations. Firstly, as Twitter's main user base is located in the U.S., investors around the world are less able to keep up with the list of opinion leaders used in this study. This geographical issue also causes an issue in regard to creating an optimal list of opinion leaders. Because of linguist limitations, Japanese opinion leaders for example were not taken into account.

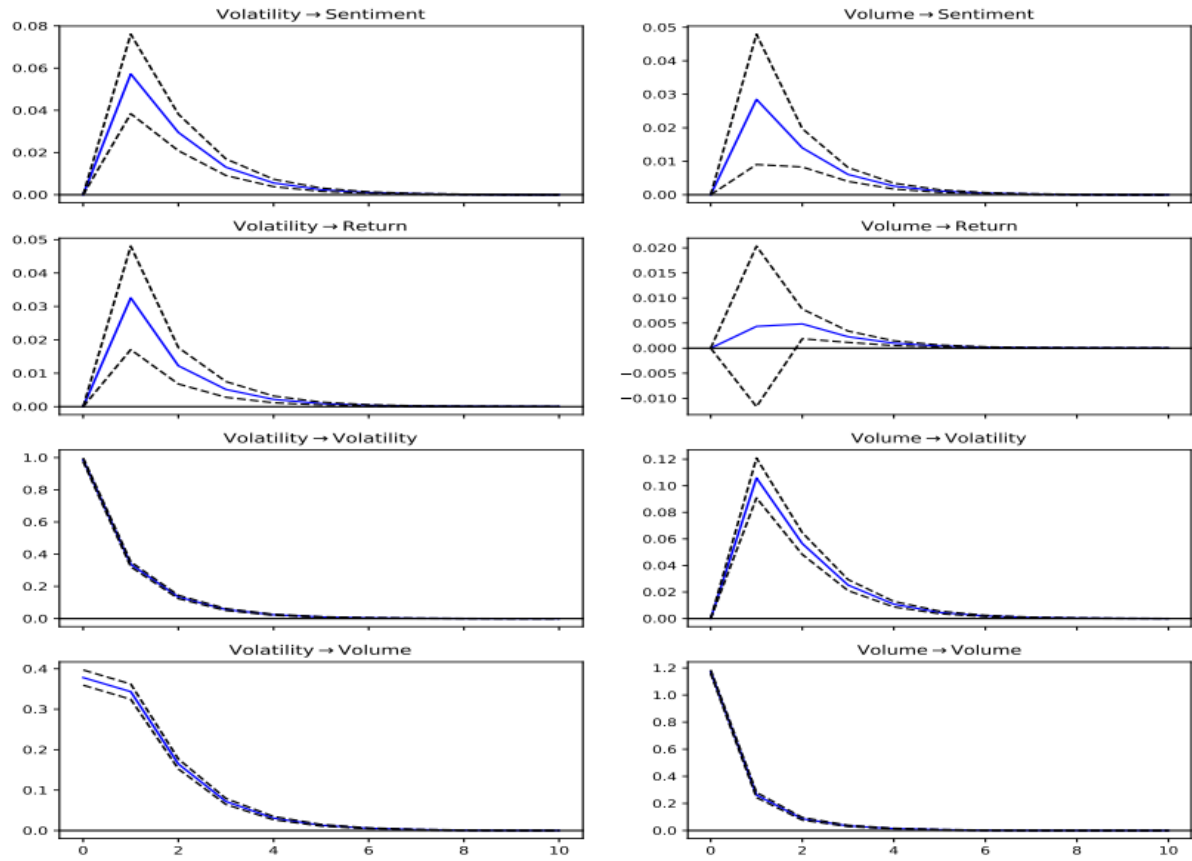
Secondly, because of the limited amount of cryptocurrency data available, only a single exchange was used to acquire the prices of Bitcoins. As noted by Makarov and Schoar [26], there exist arbitrage opportunities between exchanges locally and also between exchanges worldwide. Furthermore, Asian exchanges have experienced relatively high premiums in the past. As Bitcoin gets more popular, receives more regulatory clarity and perhaps achieves adoption there will be more price data as well as twitter data available. Additionally, this research could be replicated for cryptocurrencies other than Bitcoin

Appendix A

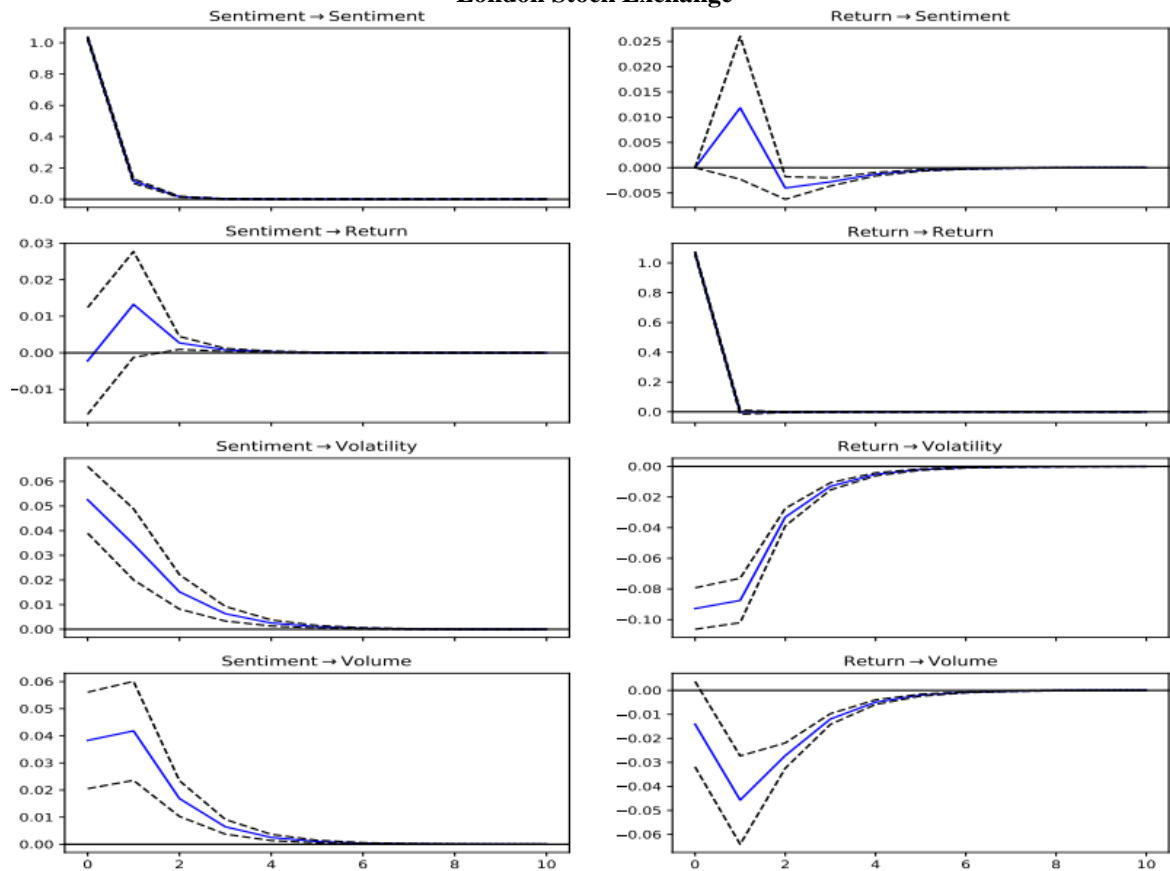
Orthogonalized Impulse response functions of 15-minute interval data

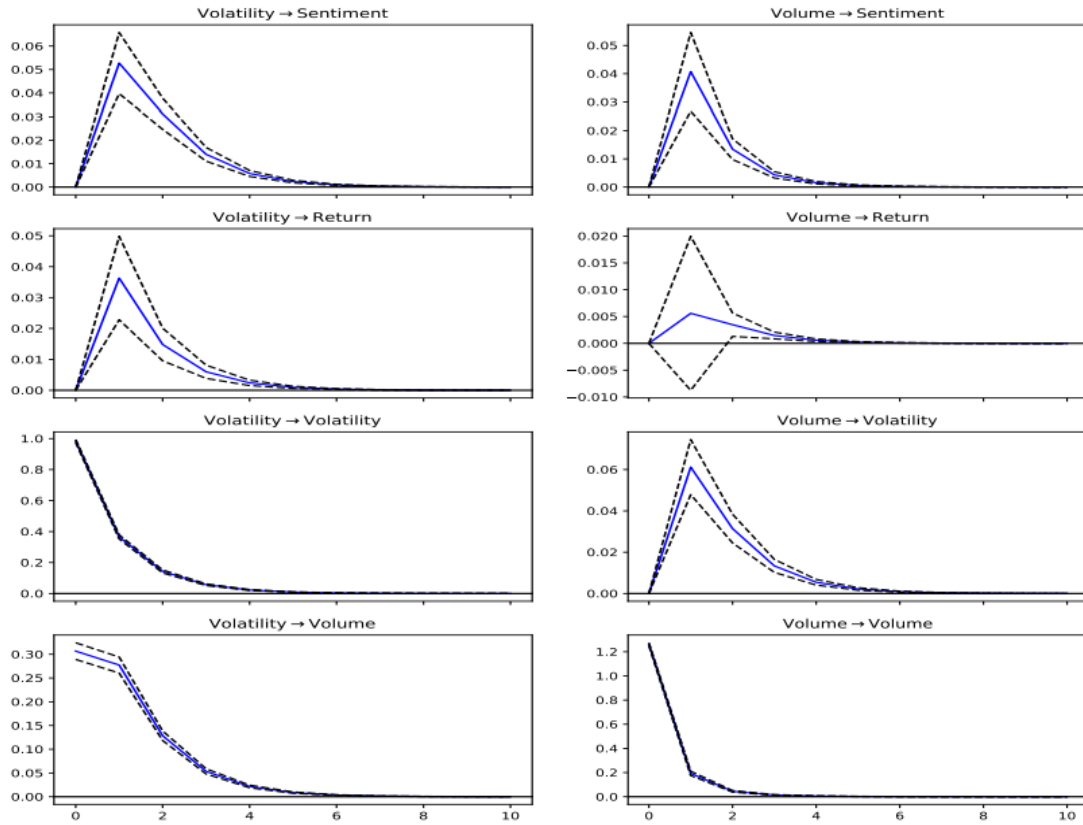
New York Stock Exchange



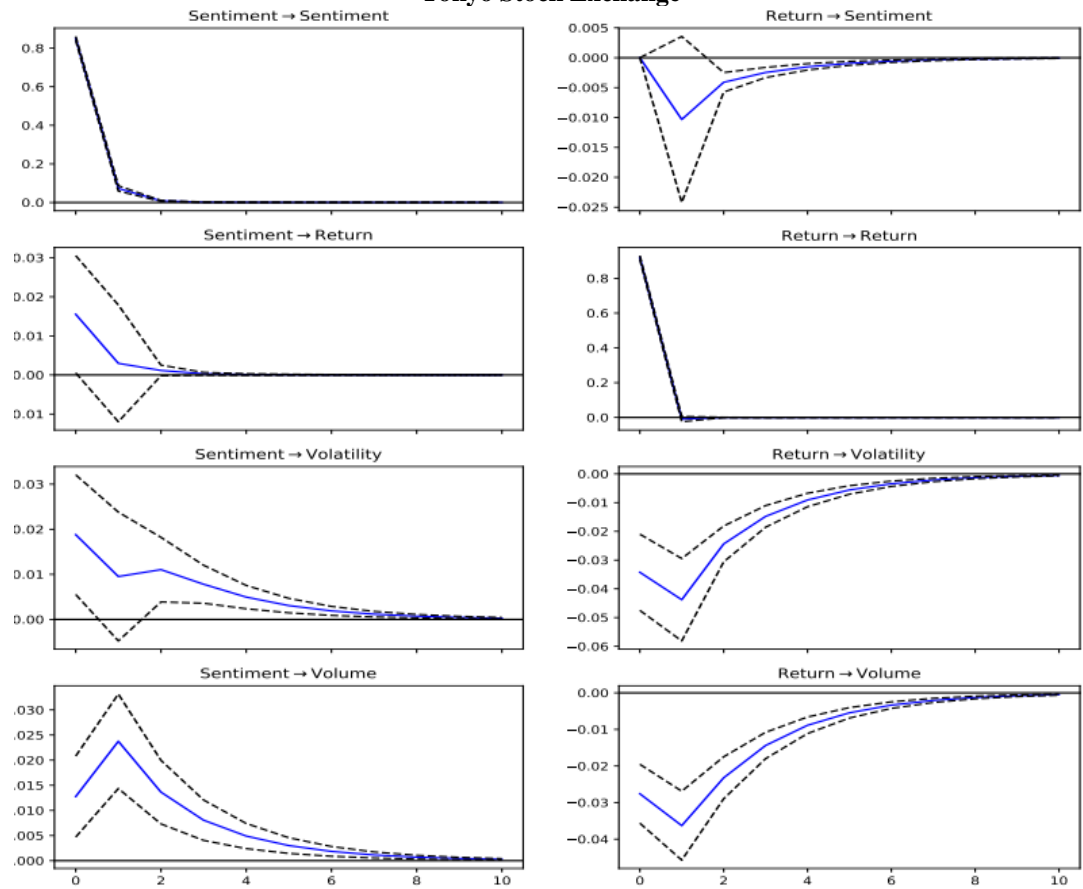


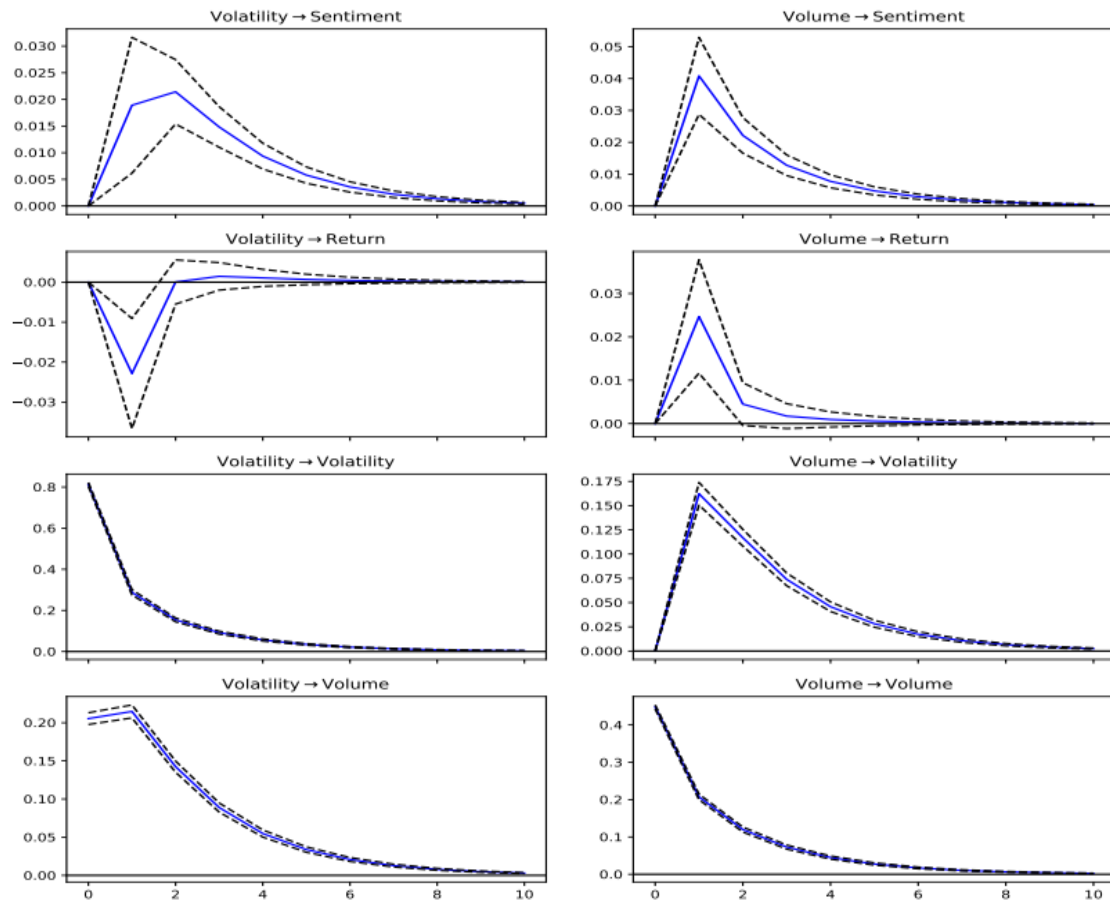
London Stock Exchange



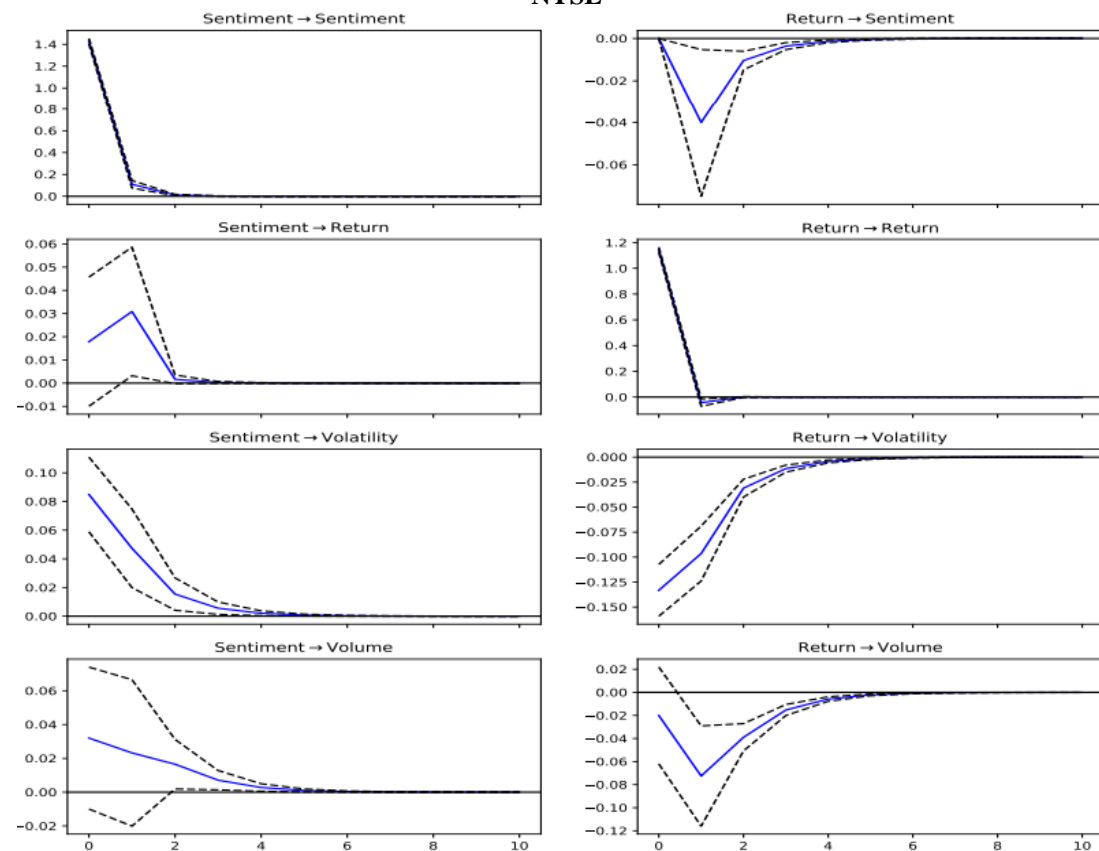


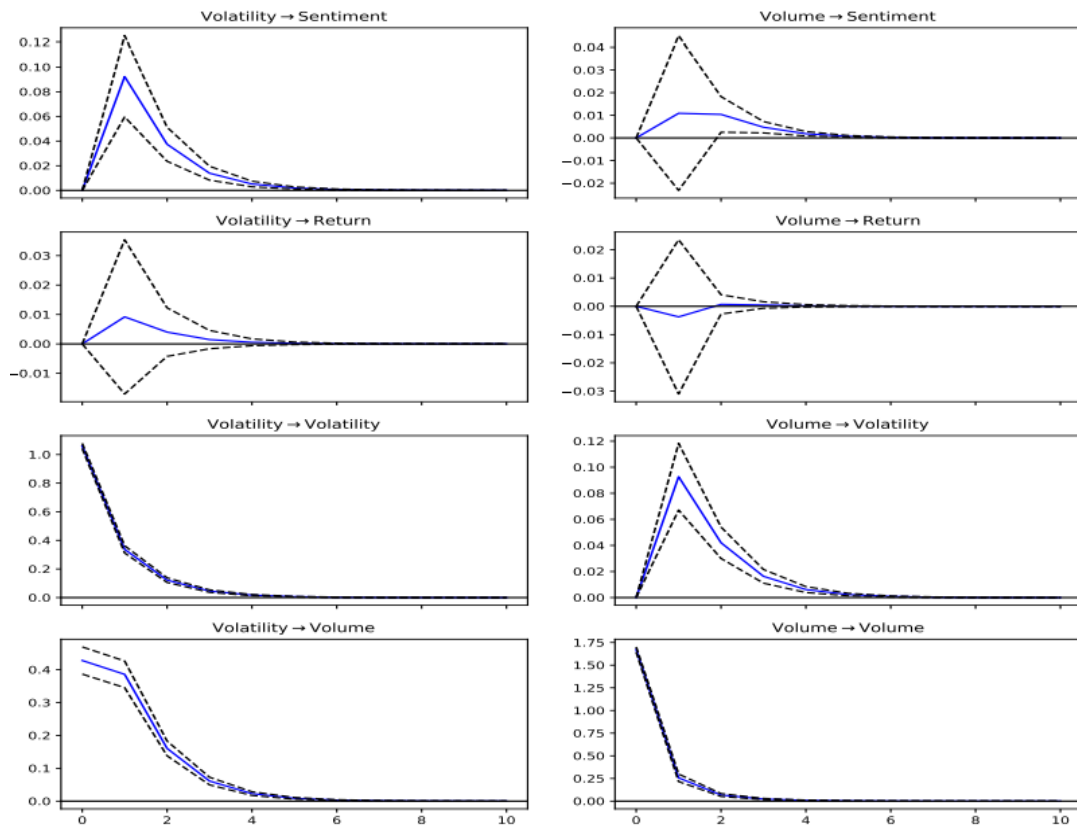
Tokyo Stock Exchange



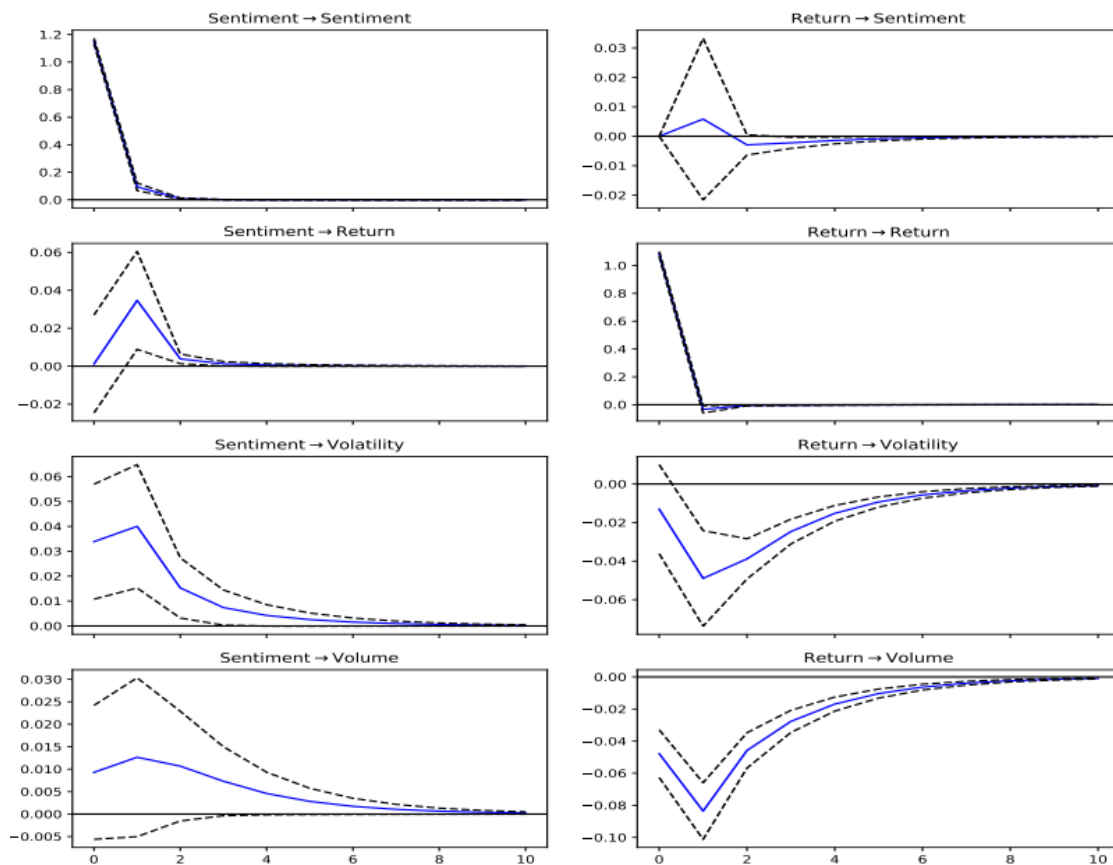


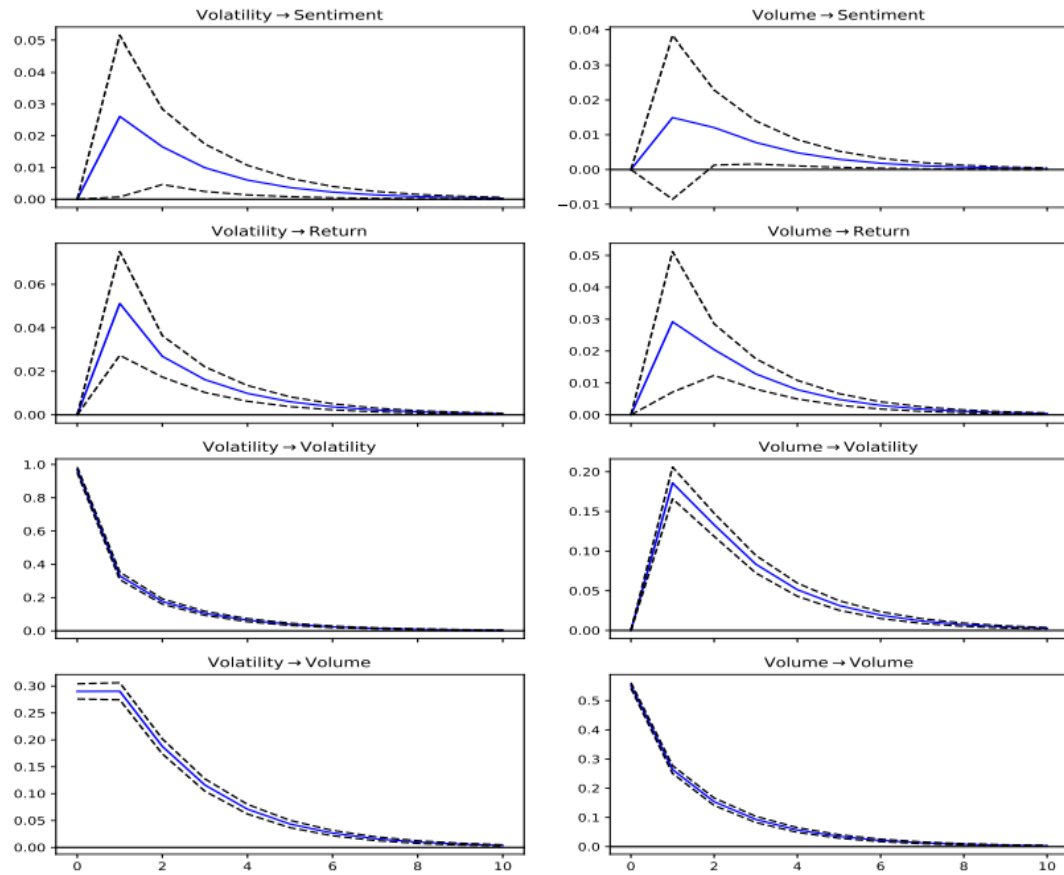
Orthogonalized Impulse response functions of 15-minute interval data in a bull market using trading hours of the NYSE



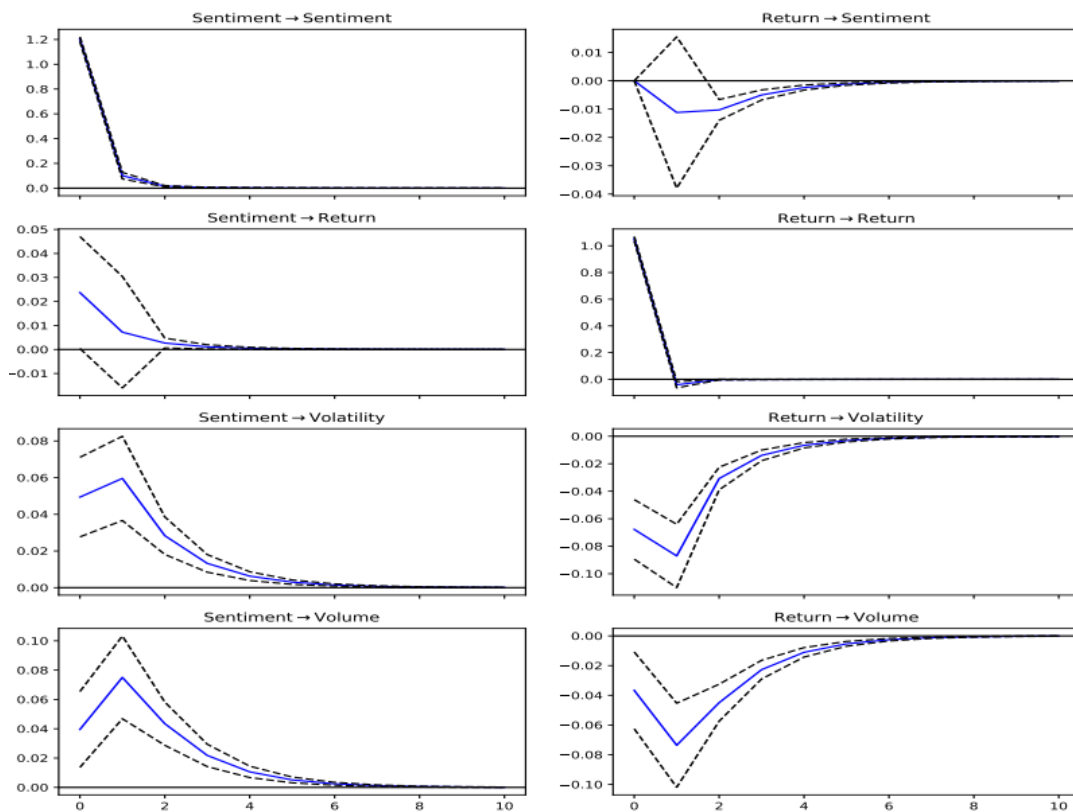


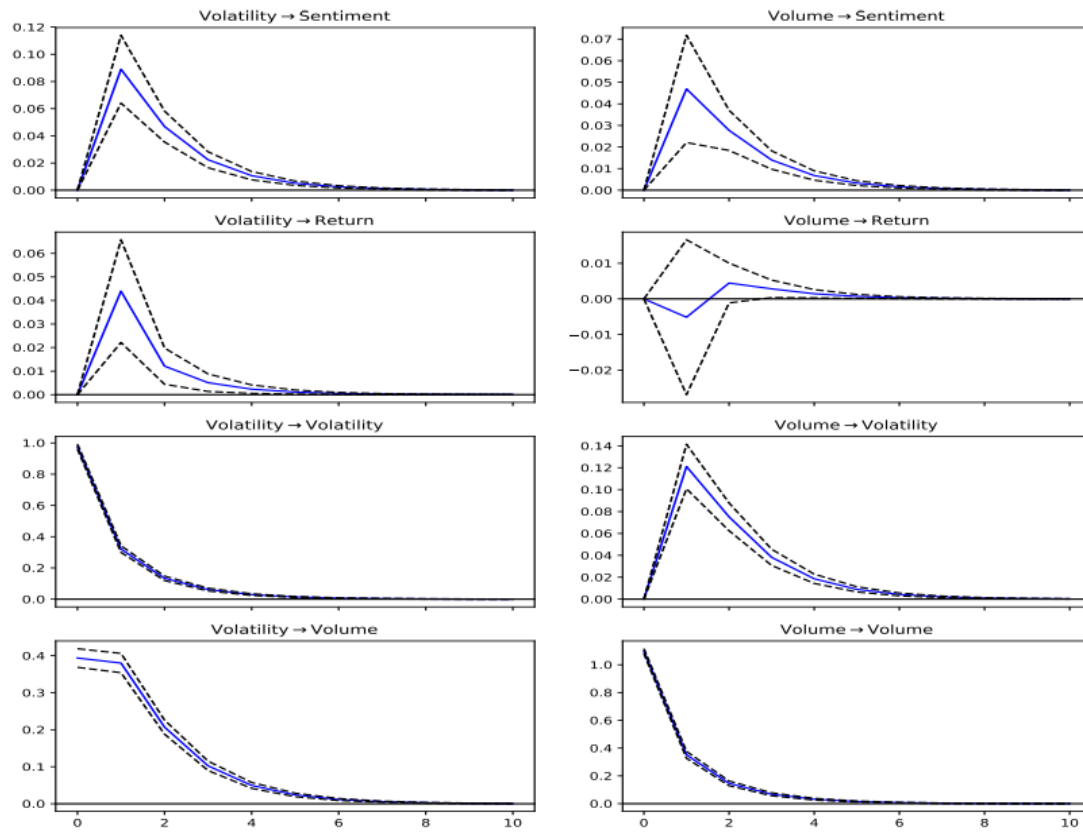
Orthogonalized Impulse response functions of 15-minute interval data in a bear market using trading hours of the NYSE



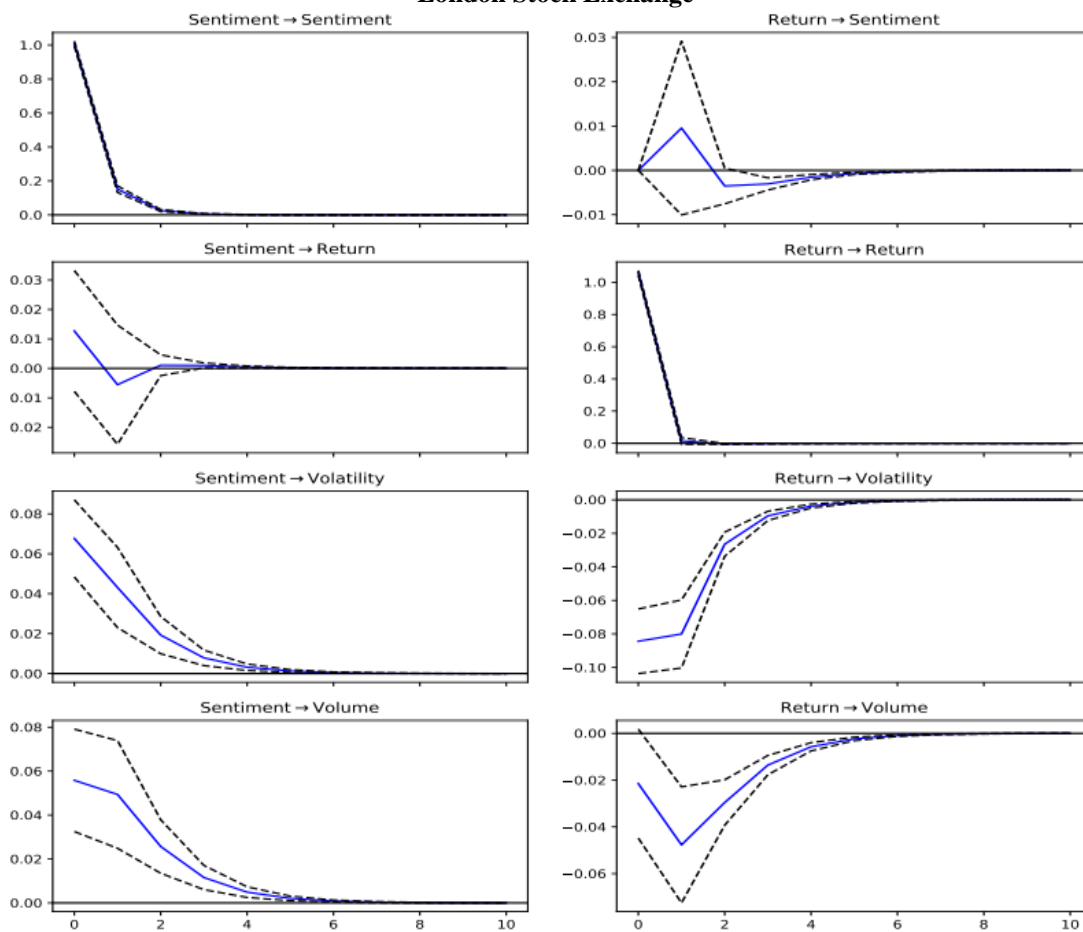


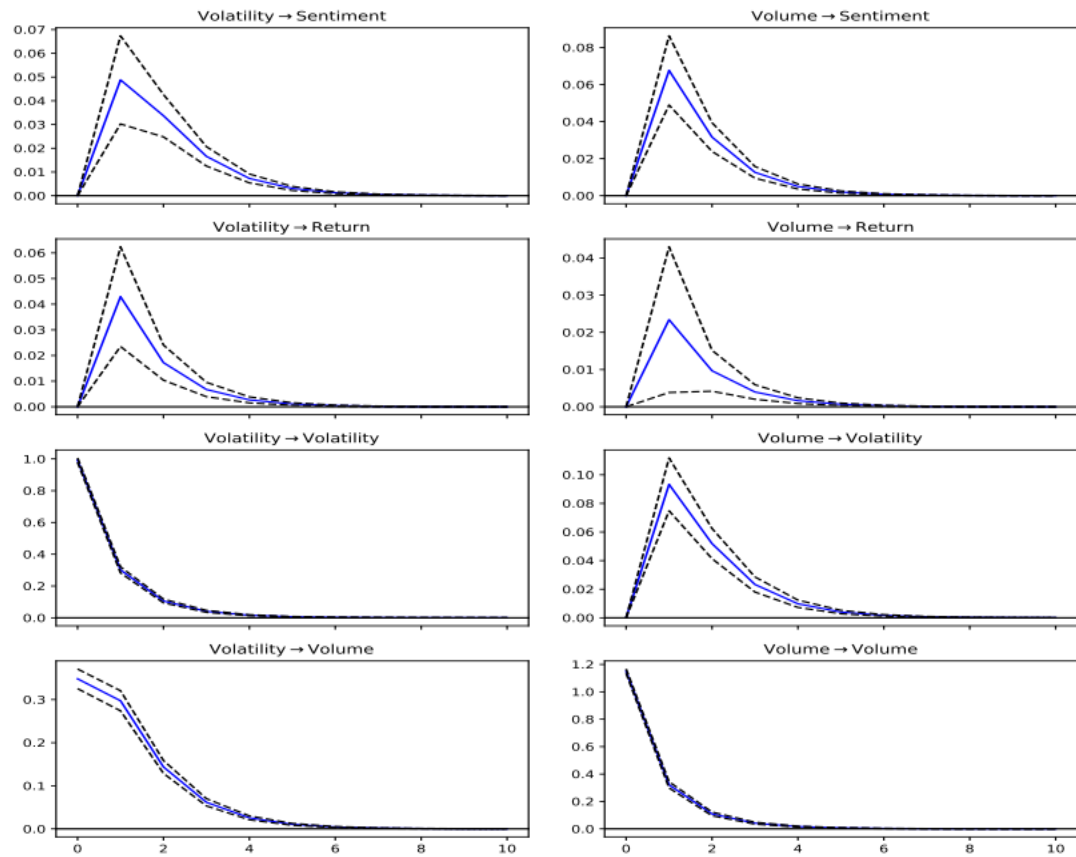
Appendix B
Orthogonalized Impulse response functions of 30-minute interval data
New York Stock Exchange



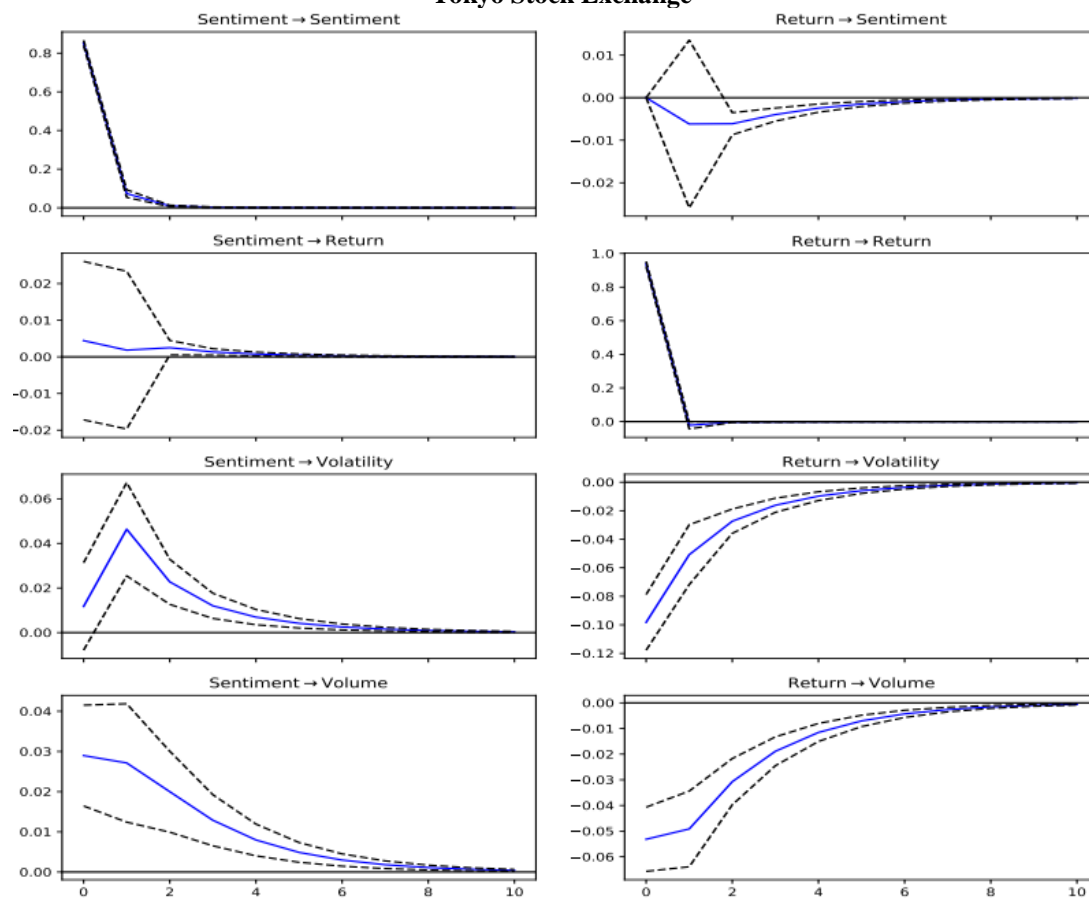


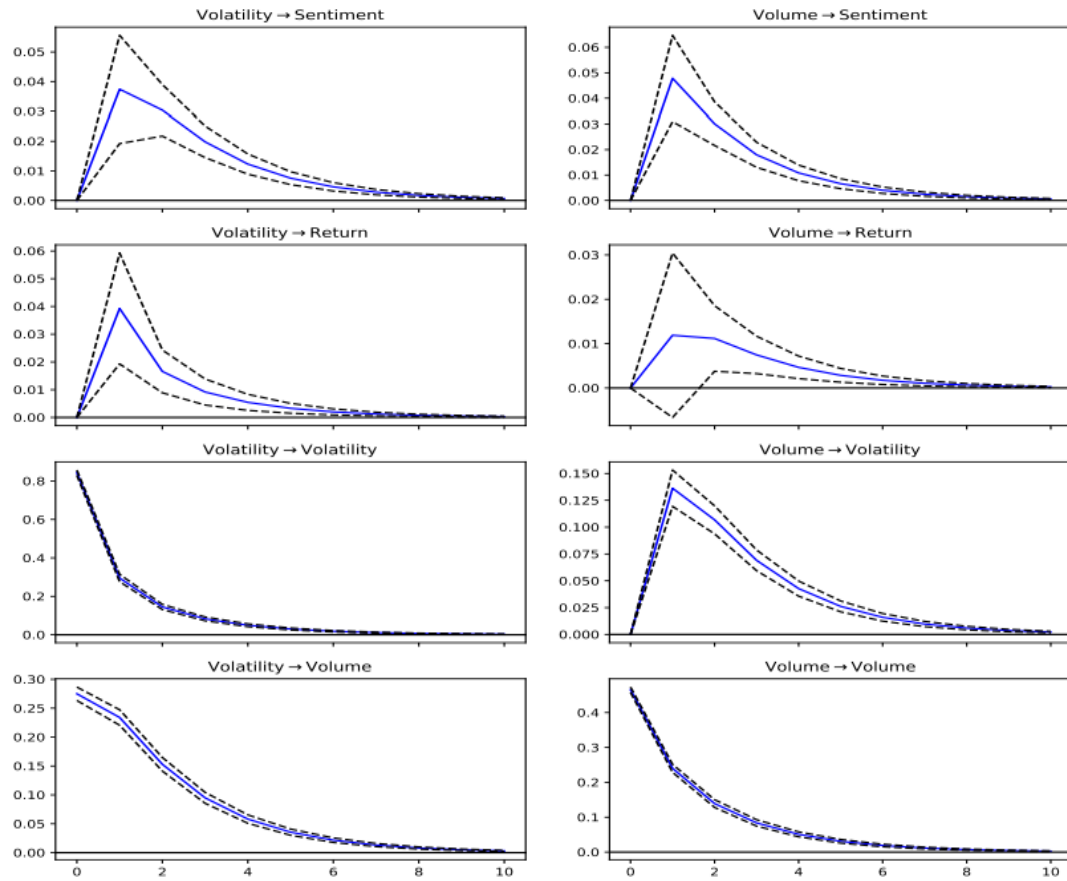
London Stock Exchange





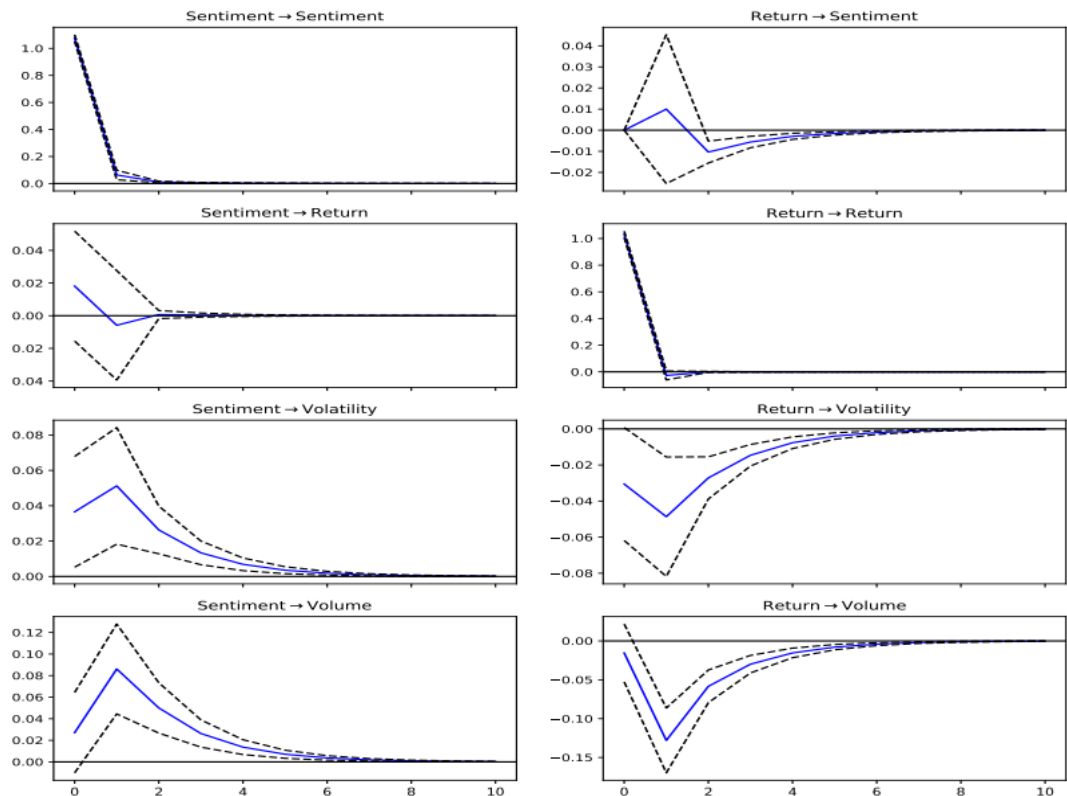
Tokyo Stock Exchange

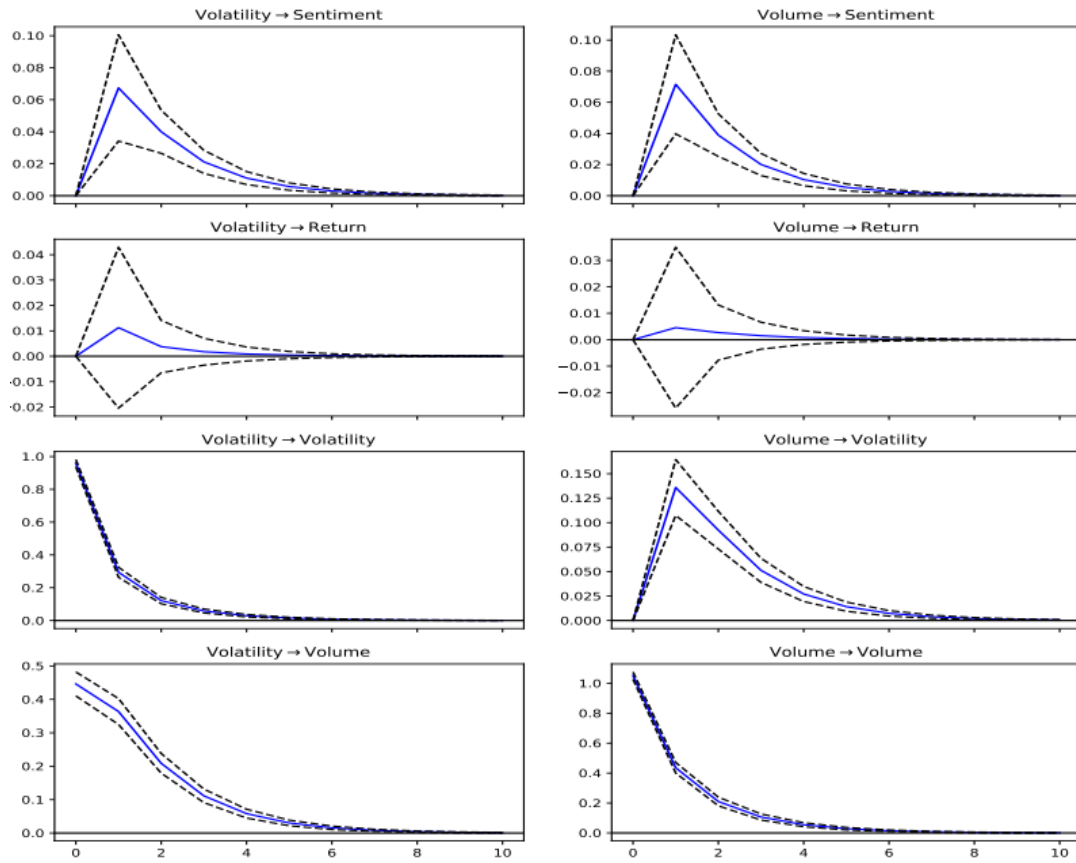




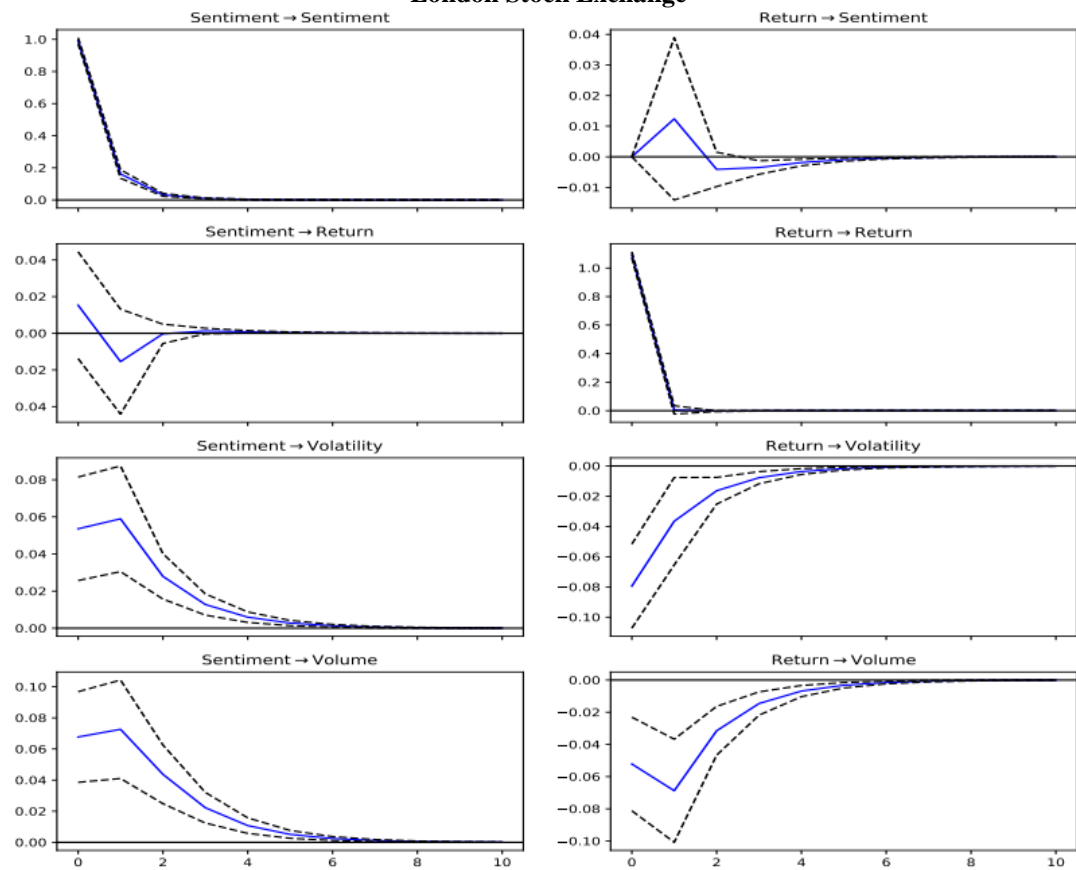
Appendix C

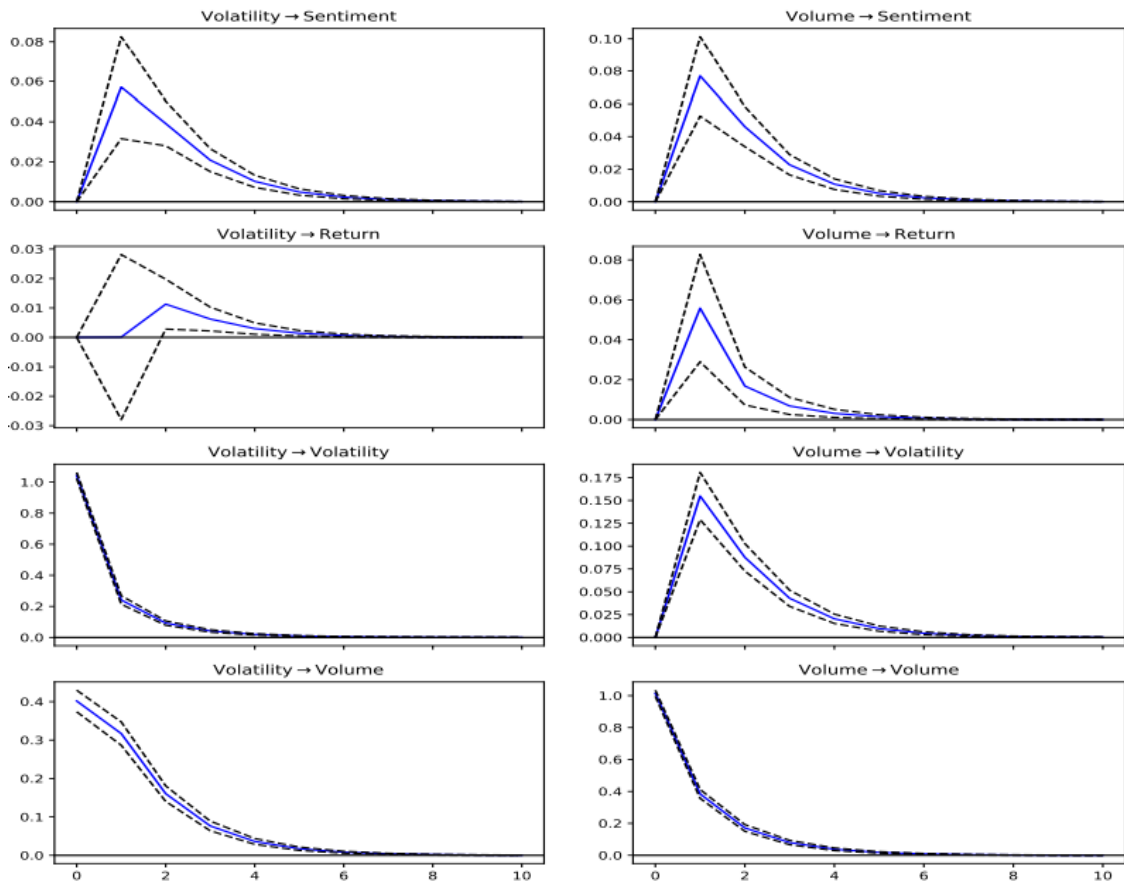
Orthogonalized Impulse response functions of 60-minute interval data
New York Stock Exchange



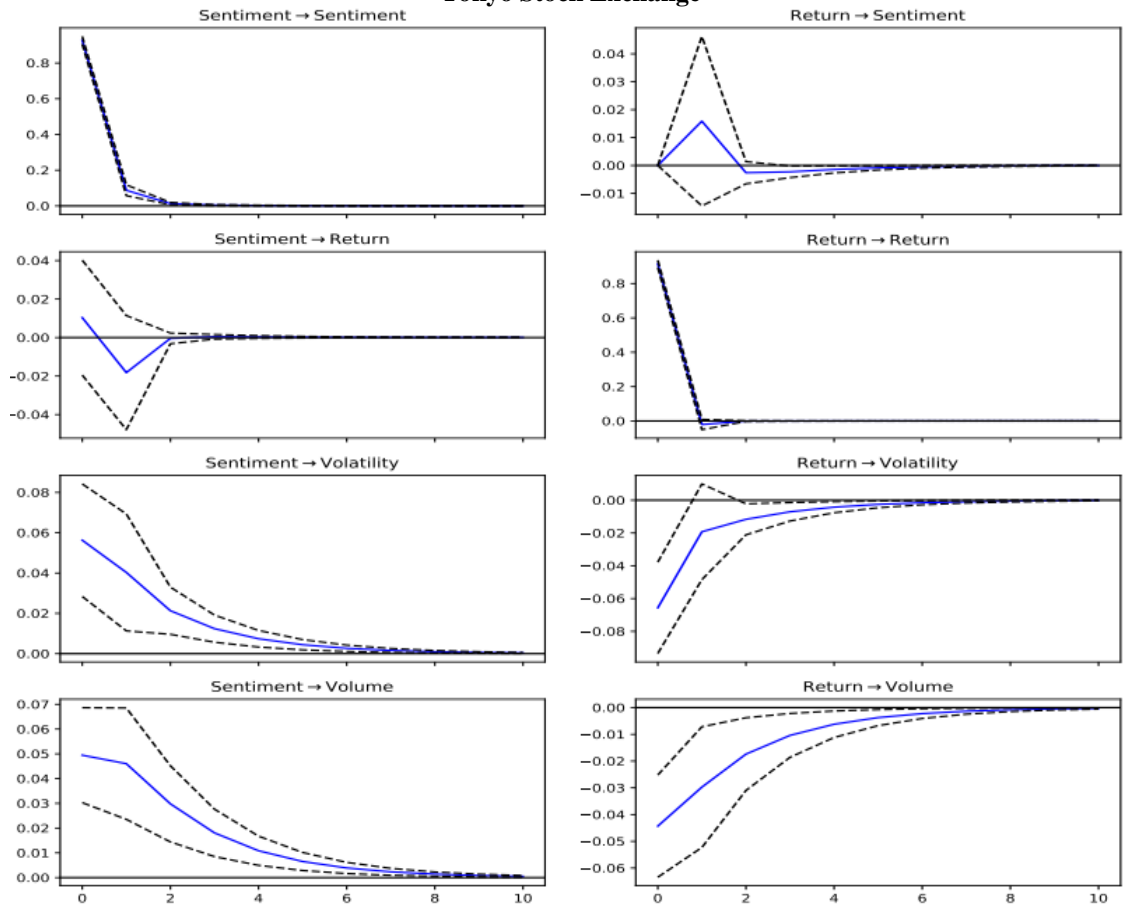


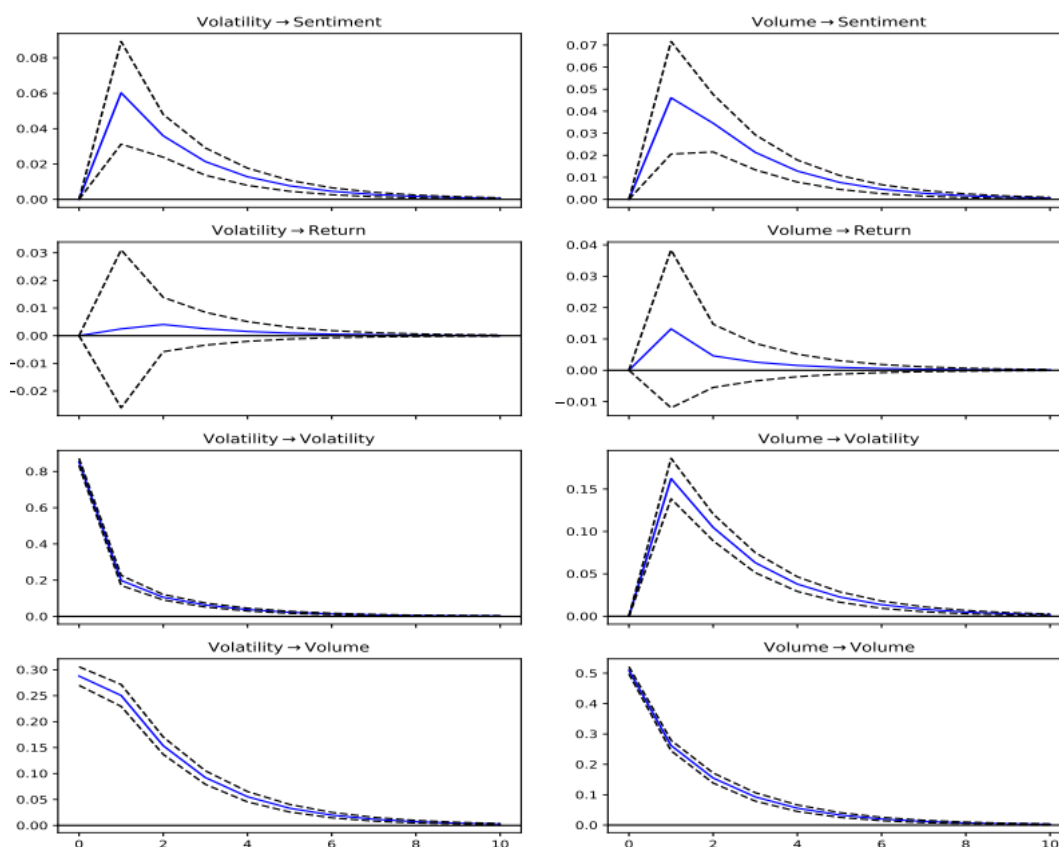
London Stock Exchange





Tokyo Stock Exchange





Acknowledgements: During the writing of this thesis I got help from my supervisor Shihao Yu.

REFERENCES

- Schuettel, P. (2017). The concise Fintech compendium. Fribourg: School of Management Fribourg/Switzerland.
- Nakamoto, S., & Bitcoin, A. (2008). A peer-to-peer electronic cash system. Bitcoin.—URL: <https://bitcoin.org/bitcoin.pdf>, 4.
- Bitcoinprice,charts,marketcap,andothermetrics—coinmarketcap. <https://coinmarketcap.com/currencies/bitcoin/#markets>, 2019. (Accessed on 05/20/2019).
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- Java, A., Song, X., Finin, T., & Tseng, B. (2007, August). Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis* (pp. 56-65).
- Fantazzini, D., Nigmatullin, E., Sukhanovskaya, V., & Ivliev, S. (2016). Everything you always wanted to know about bitcoin modelling but were afraid to ask. *Applied Econometrics*, Forthcoming.
- Liu, Y., & Tsyvinski, A. (2018). Risks and returns of cryptocurrency (No. w24877). National Bureau of Economic Research.
- Cheah, E. T., & Fry, J. (2015). Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics Letters*, 130, 32-36.
- Number of social media users worldwide 2010-2021—statista. <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/> 2017. (Accessed on 05/20/2019).
- Christina, N. (2019). 130+ social media statistics that matter to marketers in 2019. <https://blog.hootsuite.com/social-media-statistics-for-social-media-managers/>, March 2019. (Accessed on 05/20/2019).
- Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (Vol. 10, No. 2010, pp. 1320-1326).
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2(1), 1-8.
- Oh, C., & Sheng, O. (2011). Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. In *Iciss*, 1-19.
- Smailović, J., Grčar, M., Lavrač, N., & Žnidaršič, M. (2014). Stream-based active learning for sentiment analysis in the financial domain. *Information sciences*, 285, 181-203.
- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of Twitter sentiment on stock price returns. *PloS one*, 10(9), e0138441.

16. Mai, F., Bai, Q., Shan, J., Wang, X. S., & Chiang, R. H. (2015). The impacts of social media on Bitcoin performance.
17. Colianni, S., Rosales, S., & Signorotti, M. (2015). Algorithmic trading of cryptocurrency based on Twitter sentiment analysis. CS229 Project, 1-5.
18. Stenqvist, E., & Lönnö, J. (2017). Predicting Bitcoin price fluctuation with Twitter sentiment analysis.
19. Abraham, J., Higdon, D., Nelson, J., & Ibarra, J. (2018). Cryptocurrency price prediction using tweet volumes and sentiment analysis. *SMU Data Science Review*, 1(3), 1.
20. Kaminski, J. (2014). Nowcasting the bitcoin market with twitter signals. arXiv preprint arXiv:1406.7577.
21. Georgoula, I., Pournarakis, D., Bilanakos, C., Sotiropoulos, D., & Giaglis, G. M. (2015). Using time-series and sentiment analysis to detect the determinants of bitcoin prices. Available at SSRN 2607167.
22. Duan, J., Zeng, J., & Luo, B. (2014, August). Identification of opinion leaders based on user clustering and sentiment analysis. In 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) (Vol. 1, pp. 377-383). IEEE.
23. Max, B. (2017). This machine turns trump tweets into planned parenthood donations.
24. Sul, H. K., Dennis, A. R., & Yuan, L. (2017). Trading on twitter: Using social media sentiment to predict stock returns. *Decision Sciences*, 48(3), 454-488.
25. Rayarel, K. (2018). The Impact of Donald Trump's Tweets on Financial Markets. S. Bougheas, Dü.
26. Makarov, I., & Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, 135(2), 293-319.
27. Baur, D. G., Hong, K., & Lee, A. D. (2018). Bitcoin: Medium of exchange or speculative assets?. *Journal of International Financial Markets, Institutions and Money*, 54, 177-189.
28. Luo, X. (2007). Consumer negative voice and firm-idiosyncratic stock returns. *Journal of marketing*, 71(3), 75-88.
29. Kouloumpis, E., Wilson, T., & Moore, J. (2011, July). Twitter sentiment analysis: The good the bad and the omg!. In Fifth International AAAI conference on weblogs and social media.
30. Kurt, W. (2019). Twitter finally shared how big its daily user base is —and it's a lot smaller than snapchat's. <https://www.vox.com/2019/2/7/18215204/twitter-daily-active-users-dau-snapchat-q4-earning> February 2019. (Accessed on 05/20/2019).
31. Twitter usage statistics internet live stats. <https://www.internetlivestats.com/twitter-statistics/2019>. (Accessed on 05/20/2019)
32. Leon, C. (2019). The top 100 crypto and blockchain influencers (on twitter). <https://medium.com/blockinfluence/the-top-100-crypto-and-blockchain-influencers-on-crypt> March 2019. (Accessed on 05/20/2019).
33. Twitter bot. <https://en.wikipedia.org/wiki/Twitterbot>, May 2019. (Accessed on 05/20/2019).
34. Jefferson, H. (2019). Github jefferson-henrique/getoldtweetspython: Aproject written inpythontogetold tweets, it bypass some limitations of twitter official api. <https://github.com/Jefferson-Henrique/GetOldTweets-python>, November 2018. (Accessed on 05/20/2019).
35. Gilbert, C. H. E., & Hutto, E. (2014, June). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) <http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf> (Vol. 81, p. 82).
36. Baur, D. G., Cahill, D., Godfrey, K., & Liu, Z. F. (2019). Bitcoin time-of-day, day-of-week and month-of-year effects in returns and trading volume. *Finance Research Letters*, 31, 78-92.
37. Zivot, E., & Wang, J. (2006). Vector autoregressive models for multivariate time series. *Modeling financial time series with S-PLUS®*, 385-429.
38. Vector autoregressions documentation. <http://www.statsmodels.org/devel/vectorar.html>, 2019. (Accessed on 05/20/2019)
39. Lütkepohl, H. (2005). New introduction to multiple time series analysis. Springer Science & Business Media.