THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF FINANCE

Examination of Calendar Effect and Sentiment Analysis on Bitcoin

CHEN-HAN LIU
SPRING 2022

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Finance
with honors in Finance

Reviewed and approved* by the following:

Brian Spangler Davis
Clinical Assistant Professor of Finance
Thesis Supervisor and Honors Adviser

Cao Charles
Smeal Chair Professor of Finance
Thesis Reader

* Electronic approvals are on file.

ABSTRACT

This research investigate if Bitcoin exhibits the calendar effect and how sentiment affects Bitcoin's pricing. I examine the day-of-the-week effect and intraday effect using regression analysis with dummy variables and power ratio analysis. Bitcoin's abnormal returns on Fridays and early mornings, as seen from regression analysis, prove the existence of these two effects, but power ratio analysis shows no anomalies between weekdays and intraday. Sentiment analysis is studied based on data that include sentiment scores and affiliated information from over 4 million posts on Twitter. The findings offer evidence that the number of tweets posted, unweighted sentiment compound score, and sentiment compound score weighted by retweets correlate more with Bitcoin's returns. The investigation also presents that the correlation of sentiment parameters such as weighted and unweighted compound scores on Bitcoin's price is not consistent over the time horizon, which forms a positive leading index, but turns negative as a simultaneous or lagged index. The degree of influence of Twitter information over Bitcoin's price is less significant 20 minutes before and after a tweet. The analysis results explain why recent research about calendar anomalies and sentiment analysis in Bitcoin's returns contradict each other. Lastly, I also note that a tweet's information correlates more with volatility than with Bitcoin returns.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my gratitude to the following people, for which I could not have completed this thesis without their tremendous support.

Dr. Brian Davis, who served as both my Thesis Supervisor and Honors Advisor. Your feedback has been crucial to the direction of this thesis, and I cannot imagine how much time you have saved me by helping me narrow the scope of my research and focus my efforts.

Dr. Quanwei Cao, who served as my Thesis Reader. Thank you for constantly reminding me about the technical requirements needed for this project. Following your suggestions gave me the best resources that were critical to the completion of this thesis.

Dr. Timothy Simin, who, despite not being a formal advisor for this thesis, shared with me your own approach on the topic and provided inspiration for me to walk my path. It was always a joy to visit you during office hours, knowing that I will emerge a better researcher from these meetings.

Dr. Chia-Nan Liu, the best father one could ever ask for and my most enthusiastic supporter. Your encouragement, especially whenever I hit a wall, was the reason I was able to finish this thesis. Thank you for helping me become the best version of myself.

Dr. Katherine Chen, the perfect mother and the most dedicated researcher I have ever known. Your instruction, comments, and criticism elevated the quality of my work. I look forward to the day when I may serve as your co-author.

Chapter 1

Information

The Efficient Market Hypothesis (EMH) proposed by Fama (1970) is the basis for examining if the returns in an asset market are efficient. The idea underlying EMH is that past information is not useful in predicting future returns; i.e., no extra gains in market return can be obtained by using already available information. It also means that the returns and the information over a period of time are independent of each other. EMH can be examined by fitting the correlation of returns for various time and incoming information. If any correlation exists, then it indicates market inefficiency.

There are many studies that measure the calendar effects using day, week, month, holiday, and other seasonality of assets. The calendar effect is defined by the price changes of an asset in a certain time period exhibiting deviation characteristics from the average, and that the deviations exist in similar time periods. Deviations in temporal differences are known as anomalies, which can be taken as an investment opportunity hence direct proof that the market is inefficient.

Among the various calendar effects, the weekend effect is one of the most studied for traditional assets, since some information are reported or obtained over a weekend when most markets are closed. It is therefore of great interest to study if weekend information has an effect on Mondays' prices. However, for the new generation of cryptocurrency markets that is traded twenty-four hours per day and seven days per week (24/7), the calendar effect goes beyond the traditional measures of days, weeks, months, and holidays. Cryptocurrency as something to buy and sell was first introduced in 2009. Due to its cost-efficiency, initial government-free characteristic, peer-to-peer online payment facility, and speculative investment nature, cryptocurrency soon drew a lot of attention among investors. One may thus expect to find differences in seasonal performance between Bitcoin and conventional assets, such as stocks, bonds, and commodities.

The arrival of massive amounts of information indicates that public available information generally arrives more frequently during business hours, when financial exchanges are open. Some researchers have integrated the demand for information, such as the active attention measure or Google search volume (GSV), as an impact indicator of asset prices. Recently, more research has linked GSV to Bitcoin price on a daily basis and found that the GSV index positively relates to Bitcoin price movements. For the characteristic of trading 24/7, one question arises: Other than the daily GSV, is there any other information, such as high frequency posts on social media, that may act as an index for Bitcoin's price movement?

The purpose of this research is to examine if EMH holds true for Bitcoin trading by answering the following issues: How the temporal anomalies vary over time and frequencies, and how information posted on social media affects Bitcoin price. The main objectives of this study can be categorized into two significant aspects: calendar effect and influence of information. The calendar effect is evaluated using linear regression analysis with dummy variables and the power ratio method. Since Bitcoin trading is 24 hours a day and 7 days a week, I examine the calendar effect from Monday through Sunday as well as during every hour within each day. The influence of information is evaluated by calculating how the Bitcoin price correlates with the sentiment scores of tweets.

The remainder of this research runs as follows. Section 2 discusses the literature review. Section 3 presents the data and the methodology. Section 4 cites the research findings. Section 5 provides the conclusion and discussion.

Chapter 2

Literature Review

With its status as the world's leading cryptocurrency by market capitalization, there has been a significant interest in the study of Bitcoin and calendar anomalies in its returns and volatility. The main benefit of researching Bitcoin's seasonality is that it helps further determine the validity of the efficiency market hypothesis. Another benefit is to allow investors to improve their investment portfolio performance. Recent studies analyzing calendar anomalies in Bitcoin include Aharon and Qadan (2019), Baur et al. (2019), Kaiser (2019), Ma and Tanizaki (2019), and Kinateder and Papavassiliou (2021), among others. The most comprehensive research method is to use a generalized autoregressive conditional heteroskedastic (GARCH) model with dummy variables. For example, Kinateder and Papavassiliou (2021) used the GJR-GARCH(1,1) model proposed by Glosten et al. (1993). The most extensively investigated form of seasonality in these papers is the day-of-the-week effect.

Caporale and Plastun (2018) examined the day-of-the-week effect by using statistical methods such as average analysis, Student's t-test, ANOVA, and regression analysis with dummy variables. They found a Monday anomaly for Bitcoin, by which returns are considerably greater than on the other days of the week. Décourt et al. (2017) examined the Monday effect in Bitcoin by using the average daily returns on Monday, compared to other days of the week. They noted that Bitcoin tends to have higher returns on Mondays. Aharon and Qadan (2019) investigated the day-of-the-week effect with daily data for 2010-2017. Their study explained that the day-of-the-week effect is present in both the returns and volatility of Bitcoin, and that Mondays are associated with higher returns and higher volatility in Bitcoin prices compared with other days of the week. Durai and Paul (2018) presented evidence for a day-of-the-week calendar anomaly in Bitcoin returns. Decourt et al. (2019) found a statistically significant difference in the average daily returns of each day, and that Tuesdays and Wednesdays have higher returns as compared to other days. Qadan et al. (2021) collected ten years of daily data and concluded that the Monday effect on Bitcoin occurs primarily in the first three weeks of a month. Susana et al. (2020) studied a three-year daily dataset of Bitcoin prices and confirmed the existence of anomalies during Thursdays, the months March and April, and at the turn of the year. Caporale and Plastun (2019) looked at the day-of-the-week effect in the cryptocurrency market by using statistical methods such as average analysis, Student's t-test, ANOVA, and regression analysis with dummy variables. They offered no conclusive evidence against market efficiency for Bitcoin, or that Mondays' returns are considerably greater than on any other day-of-the-week. In conclusion, the existence of calendar anomalies is not consistent with the Efficient Market Hypothesis (EMH).

Yaya and Ogbonna (2019) and Kinateder and Papavassiliou (2021) conversely found no significant proof of the day-of-the-week effect in Bitcoin returns that support market efficiency. This validates the view that Bitcoin returns exhibit mostly weak-form efficiency with respect to calendar anomalies, which is in line with the findings of Nadarajah and Chu (2017) and Baur et al. (2019). The absence of significant calendar anomalies indicates that there are no seasonal patterns in returns that could be used by investors to generate abnormal returns, based on past Bitcoin price information. Kurihara and Fukushima (2017) examined the day effect in Bitcoin trading. They concluded that Bitcoin shows anomalies on weekends during the earlier period when Bitcoin was not as widely known nor traded, but Bitcoin trading has now become more efficient, and its returns should be random in the future.

Onecrucial obstacle for the financial market in general is that investor sentiment is not directly observable. Different proxies have been used, such as analyst opinions and articles in newspaper (Sadka and Scherbina 2007). Recently, the availability of extensive online discussions can be helpful to analyze individuals' statements and opinions. Financial sentiment analyses have been conducted on information over the Internet to understand the effect of reactions and emotions on the financial market. Sentiment analysis is used to evaluate the emotion of investors to understand their attitudes and thoughts that affect trading behavior. For example, Antweiler and Frank (2004) investigated online posts on Yahoo! Finance and Raging Bull to predict market volatility and asset returns. Vega (2006) found that asset returns and volatility positively correlate with the number of analysts and media coverage and explained how it could be the case that both factors increase in order to meet the rise in information demand.

Aharon and Qadan (2019) employed an active attention measure, Google search volume (GSV), to check the attention paid to Bitcoin on the Internet on daily basis. Baig et al. (2019) also employed GSV as a measure of investor sentiment and reported a strong positive relationship between GSV and Bitcoin price clustering. Lyócsa et al. (2020) used GSV activity as a gauge of panic sentiment and found that excess search volume represents a timely and valuable data source for forecasting stock price variation. For a statistical analysis of the relationship between investor sentiment and Bitcoin returns, Kapar and Olmo (2021) used weekly data between July 2010 and May 2019 and found that financial variables such as S&P 500 or the Federal Reserve financial stress index are not statistically significant to the dynamics of Bitcoin. However, the feedback effects of individual online interest (they used Google searches) are the only power variable for Bitcoin dynamics. Da et al. (2015) constructed a Financial and Economic Attitudes Revealed by Search (FEARS) index as a new measure of investor sentiment based on daily Internet search volume. Naeem et al. (2021) compared the predictability of FEARS and Twitter Happiness sentiment on Bitcoin returns and noted that online investor sentiment is a significant predictor, and that Twitter is superior to Google as an online investor sentiment proxy, because of the nature of cryptocurrency participants who are typically young individuals and computer enthusiasts. AlNemer et al. (2021) and Anamika et al. (2021) offered similar research that used the monthly Sentix Investor Confidence index, which is based on a monthly online survey of 1600 financial analysts and institutional investors, to analyze investors' emotions. Interestingly, they found that a negative sentiment score usually indicates rising cryptocurrency prices.

Research has also been conducted to explore the relationship between investor sentiment on social media and Bitcoin price. For example, Kim and Kim (2014) used Internet messages posted on Yahoo! Finance to measure investor sentiment. Renault (2017) analyzed the messages published on StockTwits to construct intraday investor sentiment indicators. Sun et al. (2021) looked at posts on Chain Node to investigate the correlation between sentiment and the cryptocurrency market. Ranasinghe and Halgamuge (2021) collected 120,000 tweets from Twitter using keywords Bitcoin and BTC during a ten-day span (12/09/2018 to 22/09/2018) for sentiment analysis. Guégan and Renault (2021) used a database that includes about one million messages crawled from StockTwits for the analysis of the correlation between investor sentiment and Bitcoin returns. In general, Twitter and StockTwits are the two of the more common social media platforms for their abundance of users and many messages related to key words such as Bitcoin, hence providing a good representation of investor sentiment. Steyn et al. (2020) found that investor sentiment and emotions derived from stock market-related tweets are significant predictors of stock market movements. A similar analysis approach has been conducted (Li et al.

2019, Mohapatra et al., 2019, Sailunaz and Alhajj, 2019, Kraaijeveld and De Smedt, 2020, Lyu et al. 2020) to study the correlation between Twitter's sentiments and cryptocurrency price changes.

Another merit for using messages posted on social media is that they are time recorded. Both Bitcoin trading and social media operate 24 hours a day and 7 days a week. The relationship between investor sentiment and Bitcoin returns for corresponding time and for different time intervals can be studied. For example, Gao et al. (2021) studied how sentiment affects Bitcoin pricing on an hourly frequency. Guégan and Renault (2021) investigated the relationship for different time intervals ranging from one minute to one day. Jain et al. (2018) attempted to predict the prices of Bitcoin two hours in advance based on the number of positive, neutral, and negative tweets accumulated every two hours.

Ordinary Least Square (OLS) regressions and Granger causality tests are commonly applied when studying the relationship between investor sentiment and Bitcoin price (Kim and Kim, 2014; Guégan and Renault, 2021; Renault 2017). A general finding is that the Bitcoin pricing mechanism can be partially revealed by sentiment found in social media. For example, Guégan, and Renault (2021) concluded that a significant relationship exists between investor sentiment and Bitcoin returns for time intervals of up to 15 minutes, while the relationship disappears as the amount of time in each interval increases. Gao et al. (2021) found that stronger bullish sentiment significantly foreshadows higher Bitcoin returns over the time range of 24 hours, while bearish and neutral financial Twitter sentiments do not. Xie (2021) analyzed hourly data and found that the sentiment and the posting of virtual investment community messages are largely driven by past market outcomes and provide limited value-relevant information for future price prediction. Öztürk and Bilgiç (2021) found that the 50 most influential accounts may provide information driving Bitcoin investors, while other Twitter accounts simply introduce some noise.

The preprocessing of Tweets and sentiment analysis are significant for providing and building an acute prediction of Twitter sentiments. A major task for big data experts is to find the optimum preprocessing strategies. For example, Pano and Kashef (2020) saw that splitting sentences and removing Twitter-specific tags or combination generally improve the correlation of sentiment scores with Bitcoin prices. Another concern in performing sentiment analysis is how to convert tweet text into a sentiment score that represents a tweet's emotion. For example, Tetlock (2007) and Jegadeesh and Wu (2013) used dictionary-based algorithms to analyze asset characteristics. Gao et al. (2021) targeted sentiment signals of tweets based on a list of positive, negative, and uncertain words according to the Loughran-McDonald finance-specific dictionary. Some well-known sentiment analysis tools are available for use. Among them, Valence Aware Dictionary and sEntiment Reasoner (VADER) is a popular choice. VADER is a lexicon-and-rule-based sentiment analysis tool that can handle words, abbreviations, slang, emoticons, and emojis commonly found in social media. It is typically much faster than machine learning algorithms, as it requires no training. Each body of text produces a vector of sentiment scores with negative, neutral, positive, and compound polarities (https://github.com/cjhutto/vaderSentiment). For example, Mohapatra et al. (2019) and Kraaijeveld and De Smedt (2020) used VADER to assign each tweet a compound sentiment score. Pano and Kashef (2020) performed VADER-based sentiment analysis of BTC tweets during the era of COVID-19.Öztürk and Bilgiç (2021) and Ibrahim (2021) also applied Valence Aware Dictionary and Sentiment Reasoner (VADER) in a logistic model to calculate positive, negative, and neutral scores.

This literature review indicates a wide array of conclusions and questions.

1. Some recent research results about calendar anomalies in Bitcoin returns contradict one another. The validation of the market efficiency of Bitcoin is still not clear.
2. The Bitcoin market is open 24/7, and few studies examine the calendar effect intraday.
3. Some results about the sentiment effect on Bitcoin returns contradict one another. Is sentiment a significant factor to Bitcoin returns? Is sentiment score a positive indicator to rising Bitcoin prices? If sentiment score and Bitcoin returns strongly correlate, then what time lag is the most significant?
4. Considering the characteristics of real-time, representativeness, and quantity, the abundant messages on social media may provide indicators for investor sentiment.
5. Other parameters surrounding a tweet are not included in the discussion of sentiment scores, such as the number of favorites, retweets, and quotes. It is thus interesting to study how these affiliated factors affect the performance of Bitcoin's price.
6. Understanding calendar anomalies and the sentiment effect is helpful in developing models that forecast Bitcoin's price movement.

Chapter 3

Data and Research Approaches

Data Descriptions

Bitcoin data (including time, high, low, close price, volume information) were downloaded from https://www.CryptoDataDownload.com. Three sets of data were gathered: dataset_1, daily data (2241 entries in total) between 2015/10/08 and 2021/11/25; dataset_2, hourly data (53769 entries in total) between 2015/10/08/13:00 and 2021/11/26/00:00; and dataset_3, minute data (44620 entries in total) between 2021/10/01/00:00 to 2021/11/01/00:00. Some data (for example, for minute data 10/11/13:50-10/11/16:10 and 10/30/21:50-10/30/22:50) are missing or contain mistakes. These data are removed during follow-up analysis.

Traditional calendar effect analysis has been performed for investigating Monday through Friday effects. However, unlike traditional stocks and commodities, the Bitcoin market is not traded at a regulated time period. The Bitcoin market is open to trade 24/7. Therefore, the calendar effect analysis is modified. The dataset_1 is used for examining the day-of-the-week (Monday through Sunday) effect, while dataset_2 is used to examine if there are abnormal returns intraday. The dataset_3 is used for tweets analysis.

The tweets posted on Twitter including Bitcoin keywords (bitcoin, bitcoins, BTC) were scraped for analysis by assuming that Bitcoin's market characteristics closely relate to these keywords. These key words cover discussions, comments, expectations, and emotions on Bitcoin's price movements. In total, 4,293,699 tweets posted between 2021/10/01/00:00 and 2021/10/31/00:00 are scraped. For each tweet, the compound sentiment score, which will be described later, is assessed using VADER. The numbers of favorites, retweets, and quotes for each tweet are also scraped. The rationale for considering such affiliated information is that some tweets receive more attention than others, and so these posts should have more weight than those that receive less attention.

Bitcoin Returns

The returns are calculated as the logarithmic value of the closing price of time t divided by the closing price of time t-1. For example, for dataset_1, the return for day (R_d) is calculated as the logarithmic value of the closing price of day t ($P_{d,t}$) divided by the closing price of the day prior to it ($P_{d,t-1}$):

$$R_{d,t} = \ln (P_{d,t}/P_{d,t-1}) \tag{1}$$

Dataset_2 aims to check if any abnormal returns exist intraday. The returns for each interval (R_h) are calculated as the logarithmic value of the closing price of hour t ($P_{h,t}$) divided by the closing price prior to it ($P_{h,t--1}$):

$$R_{h,t} = \ln (P_{h,t}/P_{h,t-1}) \tag{2}$$

Dataset_3 is used for tweets analysis. The minute data are compiled into 10-minute intervals. The scraped Bitcoin-related tweets data corresponding to the intervals are used to evaluate the relationship between tweets and Bitcoin returns. The returns for each interval (R_m) are calculated as the logarithmic value of the closing price of minute t ($P_{m,t}$) divided by the closing price 10 minutes prior to it ($P_{m,t-10}$):

$$R_{m,t} = \ln (P_{m,t}/P_{m,t-10}) \tag{3}$$

Sentiment Analysis of Tweets

A crucial aspect of this research is to test how information has an effect on Bitcoin prices. Traditional information sources such as newspaper, television, and other broadcast media may not be appropriate, considering the 24/7 online trading characteristic of Bitcoin. The availability of extensive online posts on social media about Bitcoin provides ample and prompt information, statements, opinions, and sentiment of potential investors. More than 4 million tweets posted in October 2021 are scraped on Twitter. This research explores the joint time-series behavior of the sentiment measures in these tweets and Bitcoin’s return and price volatility. Using VADER, tweet text is converted into a sentiment score that is representative to its emotion(Hutto and Gilbert 2014).

A sentiment lexicon is a mapping from tokens (words, stems of words, abbreviations, etc.) to a numerical indicator of sentiment. Each token carries a certain valence (negative, neutral, or positive sentiment) irrespective of context. VADER employs simple rules to improve its sentiment ratings for whole sentences with a corresponding valence between -1 (very negative) and 1 (very positive). VADER is an appropriate sentiment analysis tool in this study’s analysis of Bitcoin for it is specifically trained for online datasets on social media and commonly applied in recent research. VADER is also embedded in Python packages.

For each tweet, the compound score, which is a weighted average of sentiment normalized to values between -1 (extremely negative) and 1 (extremely positive), is assessed using VADER. The compound scores of the tweets over a span of 10 minutes (usually more than 500 tweets) are compiled, and the mean value is set as the unweighted sentiment index. The standard deviation of the tweets over a span is also calculated as a proxy for disagreement. The numbers of favorites, retweets, and quotes for each Tweet are also scraped. The compound sentiment is weighted by multiplying the numbers of favorites, retweets, or quotes. The weighted compound scores of the tweets over a span of 10 minutes are also compiled, and the mean value is set as the weighted sentiment indices. These unweighted and weighted indices are then used to identify correlations with Bitcoin returns from dataset_3.

Regression Analysis

Dataset_1 is used for examining the day-of-the-week effect by using the Fama-MacBeth regression format with dummy variables, as follows:

$$R_{d,t} = c_1*D_{1,t} + c_2*D_{2,t} + c_3*D_{3,t} + c_4*D_{4,t} + c_5*D_{5,t} + c_6*D_{6,t} + c_7*D_{7,t} + \varepsilon_d \quad (4)$$

Here, dummy variables $D_{i,t}$ (i=1 to 7) are 1 corresponding to Sunday, Monday, …Saturday, respectively, and otherwise 0. The random process ε_d is the normal deviate, distributed normally with mean value of 0 and variance of 1. The coefficients $c_1 \ldots c_7$ are best estimated based on ordinary least square (OLS), which provides information on the day-of-the-week anomaly of Bitcoin returns.

Dataset_2 is used for examining the intraday effect. Similarly, the regression equation with dummy variables is:

$$R_{h,t} = d_1*H_{1,t} + d_2*H_{2,t} + d_3*H_{3,t} + \ldots + d_{10}*H_{10,t} + d_{11}*H_{11,t} + d_{12}*H_{12,t} + \varepsilon_h \quad (5)$$

Here, dummy variables $H_{i,t}$ (i=1 to 12) are 1 corresponding to 0:00, 2:00, 4.00,…20:00, 24:00, respectively, and otherwise 0. The random process ε_h is the normal deviation distributed normally with a mean value of 0 and variance of 1. The coefficients $d_1 \ldots d_{12}$ are best estimated based on ordinary least square (OLS), which provides information on the intraday anomaly of Bitcoin returns.

Power Ratio Method

This research applies the power ratio model, also used in Gu (2004), to look into the calendar effect of Bitcoin. The power ratio model was originally designed to study day-of-the-week effects of the stock market. However, for the Bitcoin market, instead of trading over a regulated time, it is open for trading 24/7. Therefore, the power ratio model is modified as follows. For examining the day-of-the-week effect:

$$\text{power ratio}_{D,\,i} = (R_{D,i}/R_W),\ i=1 \text{ to } 7 \tag{6}$$

$$R_{D,i} = (1+ \text{mean return of day } i)^7 \tag{7}$$

$$R_W = (1+ \text{mean return of week}) \tag{8}$$

For example, the average value of all Monday returns for a certain year (say, 2020) is calculated and inserted into Equation (7) for $R_{D,Monday}$, while the average value of weekly returns is calculated and inserted into Equation (8) for R_W in that year. The power ratio for Monday in 2020 can then be calculated according to Equation (6). After making similar calculations, the power ratio for each day of the week in 2020 can be calculated. When the power ratio is greater than 1, the return of that day is higher than the average of the returns of the other days within the same week, and vice versa. The ratio indicates no anomaly in the return of the day if the power ratio is equal to 1.

For examining the intraday effect:

$$\text{power ratio}_{H,\,i} = (R_{H,i}/R_D),\ i=1 \text{ to } 12 \tag{9}$$

$$R_{H,i} = (1+ \text{mean return of hour } i)^{12} \tag{10}$$

$$R_D = (1+ \text{mean return of day}) \tag{11}$$

Chapter 4

Analysis Results

Figure 4-1 presents the time series of Bitcoin prices between 2015/10/9 and 2021/10/9 (period of dataset_1 and dataset _2). The time series exhibit substantial variation, covering both a distinct boom and bust cycle. This period is long enough to cover the time from when Bitcoin began to receive public attention, to its drastic price rise, and then to its sharp up and down cycles. It also covers the COVID-19 pandemic. Therefore, it is reasonable to use this dataset to determine the calendar effect.

Regression Analysis Results for the Calendar Effect

Tables 4-1 and 4-2 present the key descriptive statistics (mean, standard deviation, skewness, and kurtosis) of Bitcoin price returns and trade volumes for different days (Monday through Sunday) and different intraday times (1:00 through 24:00), respectively. The returns of Bitcoin's price were calculated using Equations (1) and (2). Furthermore, it is of interest to investigate the calendar effect for different years. Therefore, the daily and intraday Bitcoin price returns are separated into Table 4-3, and Table 4-1, respectively. The data are also illustrated in Figures 4-1 and 4-2. One approximate trend observed is that most days and intraday times have positive returns and that the returns are greater than some negative returns. This observation is clear and reasonable, because the Bitcoin price surged drastically in this period. Rigorous analyses including regression analysis and power ratio analysis using the organized data are still necessary for examining the calendar effect on Bitcoin returns.

The regression analysis with dummy variables as shown in Equation (4) is applied for day-of-the-week effect examination. The coefficients $c_1 \ldots c_7$, which corresponding to Monday through Sunday as shown in Table 4-5, are best estimated based on ordinary least square (OLS). The key descriptive statistics (including mean, standard deviation, skewness, and kurtosis) are also presented. The analysis results indicate that only Fridays' returns show the presence of day-of-the-week effect, which is statistically significant at the 1% level. The dummy's positive coefficient implies that investors earn abnormal returns on Friday. No anomaly is detected in the remaining days.

To study the intraday effect, dataset_2 is used in the analysis while the 24 hours are divided into 12 intervals of 2 hours each to reduce the number of parameters. The regression analysis with dummy variables as shown in Equation (5) is applied, and the coefficients $d_1 \ldots d_{12}$, which correspond to different time periods intraday, are shown in Table 4-5.The analysis results indicate that 10:00-12:00 (d_6), 14:00-16:00 (d_8), and 20:00-22:00 (d_{11}) possess the calendar effect on intraday returns and are statistically significant at the 10%, 5%, and 1% levels, respectively. The dummy variable's positive coefficients of these parameters imply that investors may earn abnormal returns more easily if they trade within this timeframe. No other anomalies are detected in other time periods.

Power Ratio Analysis Results for the Time Effect

Power ratio analysis of the day-of-the-week effect on Bitcoin returns for different years between 2015 and 2021 is performed using Equations (6)-(8). It is noted that the data in 2015 and in 2021 are not complete yearly data. The analysis results are tabulated in Table 4-7 and plotted in Figure 4-4. It is observed that in certain days of the week, the power ratios across different years are different and could be greater or less than 1.0. This indicates that the return of a certain day is higher or lower than the average of the returns of all other days. However, by examining the average power ratio for each day (Figure 4-4), it is observed that Friday has the highest power ratio. This is consistent with the finding of the previous described regression analysis results. ANOVA analysis of the data is also performed. The p-value is 0.12, which is not strong enough to support the hypothesis that there is an anomaly between different days within the same week.

The power ratio analysis of the intraday effect on Bitcoin returns for different years between 2015 and 2021 is performed using Equations (9)-(11). The results are tabulated in Table 4-8 and plotted in Figure 4-5. It is observed that in a certain time, the power ratios across different years vary and could be greater or less than 1.0. In the early morning, 0-4 am, the power ratios are greater than 1.0 across 2015 to 2021. The analysis' results are not consistent with the findings of the previous described regression analysis results, in that 0:00-2:00 (d_1) and 2:00-4:00 (d_2) possess a negative effect on intraday returns. ANOVA of the data is also performed. The p-value is 0.36, which is not strong enough to support the hypothesis that there is an anomaly intraday.

Correlation between Tweet Indices and Bitcoin Returns

A database containing tweets' information and dataset_3 is prepared for correlation analysis of tweets/Bitcoin returns. The duration for analysis is between 2021/10/01/00:00 and 2021/10/31/00:00. The data are compiled into 10-minute intervals. The 9 parameters crawled from the tweets include the following parameters: number of tweets (N), average sentiment compound score (SC_{avg}), average sentiment compound score weighted by favorite counts ($SC*F\_{avg}$), average sentiment compound score weighted by quote counts ($SC*Q\_{avg}$), average sentiment compound score weighted by retweet counts ($SC*RT\_{avg}$), standard deviation of sentiment compound score (SC_{sdv}), standard deviation of sentiment compound score weighted by favorite counts ($SC*F\_{sdv}$), standard deviation of sentiment compound score weighted by quote counts ($SC*Q\_{sdv}$), and standard deviation of sentiment compound score weighted by retweet counts ($SC*RT\_{sdv}$). The corresponding return (calculated using Equation (3)) and trade volume (Vol.) for each interval are appended into the database.

Table 4-10 provides summary statistics (mean, standard deviation, maximum, minimum, as well as the skewness and kurtosis) for the main variables of interest: numbers of tweets posted along with favorites, quotes, and retweet as well as sentiment scores evaluated using VADER. The subscript avg of variables indicates the average value, while the subscript sdv indicates the standard deviation of these variables in 10-minute span. As observed from the maximum and minimum values and the standard deviation, the numbers of tweets posted along with favorites, quotes, and retweets per 10 minutes are noisy. The average sentiment and the standard deviation of sentiment are also relatively noisy.

The statistics of Bitcoin return and trading volume are listed in Table 4-10.The mean return for Bitcoin (R_{avg}) within a 10-minute span is calculated using Equation (3). The mean returns for Bitcoin per 10 minutes are positive (0.0001) with a standard deviation of 0.0029. This means the fluctuation of Bitcoin returns per 10 minutes is drastically high. The average and standard deviation of each variable within a 10-minute span are calculated similarly. For example, over the time span 10/1 00:00:00 ~ 10/1 00:09:59, there are 696 posts (Table 9). Therefore, the average and standard deviation of 696 SC*F are calculated as $SC*F\_{avg}$ and $SC*F\_{sdv}$.

This database, as shown in Table 4-9, is used to identify correlations between Twitter information and Bitcoin returns. Our empirical approach extracts a sentiment measure from tweets concerning Bitcoin. In particular, we take the sentiment measure as being relative to some time-variant-based level of expectations. It is of interest to understand if these parameters are leading or lagging indicators to Bitcoin returns. Therefore, the correlation between each parameter and Bitcoin returns is evaluated for different time lags. A +10 lag indicates the parameter value is 10 minutes ahead of the return, while a -10 lag indicates the parameter value is 10 minutes behind the return, and 0 indicates the parameter value is simultaneous with the return. The investigated time lags are -30, -20, -10, 0, +10,+20, and +30 minutes.

Tables 4-11 to 4-19 list the correlation analysis results of the 9 tweet parameters against Bitcoin returns. Table 4-20 lists the correlation analysis results of trading volume against Bitcoin returns. These tables reports the correlation coefficient, standard deviation, t-statistics, p-value, and coefficient of determination (R^2). A positive correlation coefficient implies that this parameter possesses a positive correlation with Bitcoin returns and vice versa. The general trends observed from these tables are that p-values are not significant and R^2values are minimal for all cases. A large p-value indicates the correlation results are not statistically significant, and a small R^2 indicates the predictive ability of the regression model is not good.

Some further findings can still be drawn. First, the number of tweets (N) positively correlates with Bitcoin price return. This is consistent with the finding of Rognone et al. (2020) that Bitcoin reacts positively to both positive and negative news, because of investor enthusiasm for Bitcoin irrespective of the sentiment from news. The correlation is significant when N is used as a simultaneous or lagging index, and its predictive ability is relatively strong (Figure 4-6).

Second, the unweighted sentiment compound score (SC_{avg}) generally negatively correlates with the return. The correlation is significant when SC_{avg} is used as a 10~20-minute lagging index, and its predictive ability is relatively strong. Though not significant, it is interesting to find that the coefficient turns from negative to positive when the lag exceeds 20 minutes (Figure 4-7). This means that Twitter sentiment, as a leading index by 20+ minutes, positively correlates with Bitcoin returns, but as a simultaneous or lagging index it has a negative correlation.

Third, the regression analysis results for the sentiment compound score weighted by favorites ($SC*F\_{avg}$) and quotes ($SC*Q\_{avg}$) are ambiguous. The correlation coefficients do not show trends, and neither p-values nor coefficients of determination support the analysis results. Only the sentiment compound scores weighted by retweets ($SC*RT\_{avg}$) have a statistically significant negative correlation with Bitcoin returns at -20 and +20 minutes. It is interesting to find that, though not significant, the correlation coefficient of SC_{avg} runs opposite to the correlation coefficients of $SC*F\_{avg}$ and $SC*Q\_{avg}$, but this opposite phenomenon is less obvious for $SC*RT\_{avg}$.

The standard deviation is an index showing disagreement of the data population. A large standard deviation among data can be a sign of fundamental uncertainty. One proposition is that when uncertainty increases, risk-averse traders require higher future returns to absorb the risk, which leads to a fall in price. Therefore, the standard deviations of the weighted and unweighted sentiment compound scores are included into analysis to examine this proposition. The analysis results are in Tables 4-16 to 4-19. Regretfully, the results are ambiguous, and the statistical analysis results are not significant enough to support nor negate the proposition. Here, SC_{adv} acts as a negative regression coefficient at a lag of -10 minutes, while it is positive at a lag of +10 minutes, while $SC*RT\_{sdv}$ acts as a positive regression coefficient at a lag of +10 minutes. The correlation analysis results of Bitcoin returns and trade volume (Vol) for different time lags are in Table 4-20, which presents that trade volume generally positively correlates with Bitcoin price returns (Figure 4-8).

A comparison of the coefficients of determination for these analyzed cases is shown in Figure 4-9. It shows that the information retrieved from Twitter, including number of tweets posted (N), unweighted sentiment compound score (SC), and sentiment compound score weighted by retweets (SC*RT), along with trade volume (Vol), correlate with Bitcoin price returns. The average coefficient of determination for the 9 tweet parameters for different time lags is shown in Figure 4-10. As the time difference exceeds more than 20 minutes, the coefficient of determination becomes lower. This indicates that Twitter information correlates more with Bitcoin prices within a 20-minute span, no matter for leading or lagging.

Correlation between Tweet Indices and Bitcoin Volatility

Volatility is a measure of how much the price of an asset has moved up or down over time. Generally, the more volatile an asset is, the riskier it is considered to be as an investment and the more potential it has to offer either higher returns or higher losses over shorter periods of time. Bitcoin is deemed as an asset of high volatility, because of various reasons, such as lack of regulation, no intrinsic value, limited supply, speculation, media information, low barrier for inventor profile, and so on. Understanding its volatility can help an investor to decide whether to trade it, own it, or just continue watching its developments.

The statistics of standard deviations for Bitcoin returns ($R_{\_sdv}$) are listed in Table 4-10.The minute price data within a 10-minute span (i.e., 10 data points) are used to calculate the standard deviation for Bitcoin returns. It is used as a proxy for the volatility of the returns within the 10-minute span. The mean standard deviation for Bitcoin per 10 minutes is 0.0007 with a standard deviation of 0.0006. Though this mean standard deviation is not large, in terms of the coefficient of variation (cov), it is 700%. It means the fluctuation of Bitcoin price per 10 minutes is drastically high.

Tables 4-21 to 4-29 list the correlation analysis results of the 9 tweet parameters against the volatility in Bitcoin returns. Table 4-30 lists the correlation analysis results of trading volume against the volatility in Bitcoin returns. These tables report the correlation coefficient, standard deviation, t-statistics, p-value, and coefficient of determination (R^2).

The analysis results reveal some clear findings. First, the number of tweets (N) positively correlates with the volatility in returns. Small p-values indicate the analysis results are significantly validated. This finding is significant whether N is used as a leading, simultaneous, or lagging index (Figure 4-11).

Second, on the contrary, the unweighted sentiment compound score (SC_{avg}) negatively correlates with the volatility in returns, and it is significant whether SC_{avg} is used as a leading, simultaneous, or lagging index (Figure 4-12). The regression analysis results for weighted sentiment compound scores ($SC*F\_{avg}$, $SC*Q\_{avg}$, and $SC*RT\_{avg}$) are ambiguous and not significant. Only the sentiment compound scores weighted by retweets ($SC*RT\_{avg}$) have a statistically significant positive correlation with Bitcoin returns at -10 and -30 minutes. It is interesting to find that, though not significant, the correlation coefficient of SC_{avg} runs opposite to the correlation coefficients of $SC*RT\_{avg}$.

The analysis results of the standard deviations of the weighted and unweighted sentiment compound scores appear in Tables 4-26 to 4-29. Relative to the vague correlation in SC_{adv} (only the negative regression coefficient at a lag of -10 minutes is significant), the regression analysis results for all standard deviations in weighted sentiment compound scores ($SC*F\_{sdv}$, $SC*Q\_{sdv}$, and $SC*RT\_{sdv}$) positively correlate with the volatility in Bitcoin returns, and the correlation is more significant as they are used as lagged or simultaneous indices. The correlation analysis results of volatility in Bitcoin returns and trade volume (Vol) for different time lags are in Table 4-30. It is found that trade volume positively correlates with volatility, and this correlation is significant no matter when used as a leading, simultaneous, or lagging index (Figure 4-13).

A comparison of the coefficients of determination for these analyzed cases is shown in Figure 4-14. It is seen that the information retrieved from Twitter, including number of tweets posted (N), unweighted sentiment compound score (SC), and standard deviation in weighted sentiment compound scores, along with trade volume (Vol), correlates more with volatility in Bitcoin price returns. The average coefficient of determination for the 9 tweet parameters for different time lags is shown in Figure 4-15. It noticeably illustrates that Twitter information mostly correlates with the volatility in Bitcoin price simultaneously and correlates less as time elapses. Comparing Figure 4-10 with Figure 4-15, it is obvious that Twitter information correlates more with volatility than with Bitcoin returns.

Table 4-1 Descriptive statistics of weekday data (Dataset_1) of Bitcoin price return and trade volume

Weekday	Daily Return				Daily Volume			
	mean	sdv	skewness	kurtosis	mean	sdv	skewness	kurtosis
Monday	0.00299	0.0440	-0.33	4.73	3662.86	3850.82	2.63	8.15
Tuesday	0.00160	0.0432	-0.13	3.59	3898.58	4222.86	3.07	11.79
Wednsday	0.00355	0.0420	-0.36	2.98	4017.87	3829.77	2.13	4.97
Thursday	0.00001	0.0511	-1.40	10.15	4294.12	5272.35	4.25	28.60
Friday	0.00677	0.0408	0.25	5.76	3757.34	4301.03	3.25	15.07
Saturday	-0.00019	0.0362	-0.87	4.50	1934.63	2376.26	2.79	9.47
Sunday	0.00264	0.0358	0.38	6.55	2111.55	2628.10	3.05	13.80

Table 4-2 Descriptive statistics of intraday data (Dataset_2) of Bitcoin price return and trade volume

O'clock	Return				Volume			
	mean	sdv	skewness	kurtosis	mean	sdv	skewness	kurtosis
0	0.00010	0.00947	-0.58	10.35	136.54	243.35	7.00	83.31
1	0.00019	0.00871	-0.80	29.62	126.64	212.81	4.54	29.71
2	-0.00033	0.00881	3.19	92.84	122.75	198.12	4.16	28.39
3	-0.00028	0.00746	-0.11	14.49	116.36	229.90	7.23	86.19
4	-0.00011	0.00826	0.22	31.14	108.83	186.98	4.85	36.83
5	-0.00012	0.00758	-0.97	14.88	99.00	181.40	5.86	54.08
6	0.00019	0.00751	-0.96	17.10	93.12	171.01	6.13	58.15
7	0.00035	0.00793	0.98	17.37	91.63	158.44	6.58	75.84
8	-0.00005	0.00868	0.55	23.31	92.53	159.95	5.38	45.16
9	0.00002	0.00840	0.52	16.60	93.54	158.23	4.01	22.85
10	-0.00006	0.00933	-4.56	81.93	99.38	182.65	4.44	26.53
11	0.00024	0.00836	0.90	19.49	141.26	284.95	5.01	42.68
12	0.00036	0.00977	0.01	26.90	140.03	336.70	11.68	203.27
13	0.00018	0.01004	-1.26	20.89	167.37	338.06	10.20	191.18
14	-0.00021	0.00973	0.10	16.93	195.92	395.72	9.16	130.48
15	0.00042	0.00908	0.83	11.73	190.76	302.93	5.59	48.50
16	0.00015	0.00997	-0.17	14.27	193.20	278.41	4.48	34.55
17	0.00002	0.00842	-0.48	14.07	178.15	337.71	10.08	160.90
18	-0.00001	0.00768	-1.19	17.46	162.74	262.50	6.39	71.35
19	0.00025	0.00828	0.41	20.40	168.63	280.82	5.90	54.23
20	-0.00006	0.00927	-0.31	15.03	258.94	442.33	3.73	18.60
21	0.00070	0.00958	-1.26	56.97	187.99	322.95	4.60	39.41
22	0.00044	0.00992	-4.31	94.40	127.60	188.39	4.04	24.64
23	0.00006	0.01015	-1.77	79.79	133.95	249.62	8.91	154.69

Table 4-3 Daily returns of Bitcoin price (Dataset_1)

Weekday	2015	2016	2017	2018	2019	2020	2021
Monday	0.02077	0.00048	0.01469	-0.01382	0.00559	0.01066	-0.00415
Tuesday	-0.00154	-0.00235	0.00881	0.00010	-0.00275	0.00539	0.00116
Wednesday	0.01501	0.00504	0.00729	-0.00807	0.00124	0.00606	0.00692
Thursday	0.00160	0.00392	0.00992	-0.00617	-0.00807	0.00001	-0.00030
Friday	0.02057	0.00414	-0.00534	0.00848	0.01705	0.00388	0.00501
Saturday	-0.00605	0.00113	0.00395	-0.00762	-0.00043	0.00271	-0.00073
Sunday	-0.00250	0.00303	0.01097	0.00257	0.00031	-0.00179	0.00260

Table 4-4 Intraday returns of Bitcoin price (Dataset_2)

O'clock	2015	2016	2017	2018	2019	2020	2021
0	0.00014	0.00008	0.00264	0.00053	0.00086	-0.00001	0.00006
2	0.00034	0.00028	0.00089	-0.00038	0.00071	0.00070	0.00006
4	0.00082	0.00057	-0.00001	-0.00096	-0.00008	-0.00034	-0.00009
6	0.00034	-0.00029	0.00084	-0.00124	-0.00034	-0.00018	-0.00003
8	0.00003	-0.00048	-0.00014	0.00068	-0.00090	0.00012	-0.00003
10	0.00132	0.00035	0.00178	-0.00082	-0.00023	-0.00047	-0.00002
12	0.00213	0.00055	0.00001	-0.00131	0.00022	-0.00108	0.00005
14	0.00320	0.00024	0.00005	0.00169	0.00012	-0.00051	-0.00006
16	-0.00177	0.00032	-0.00016	-0.00068	0.00005	-0.00007	0.00001
18	0.00044	-0.00027	0.00163	-0.00010	0.00103	-0.00144	0.00001
20	0.00016	0.00092	-0.00036	-0.00072	0.00013	-0.00149	-0.00010
22	-0.00051	0.00004	-0.00001	-0.00028	0.00020	0.00096	-0.00005

Table 4-5 Regression analysis results of weekday effect on Bitcoin price return

parameter	coefficient	sdv	t-statistic	p- value
c_1	0.00296	0.00234	1.266	0.206
c_2	0.00160	0.00234	0.683	0.495
c_3	0.00349	0.00234	1.490	0.136
c_4	-0.00005	0.00234	-0.020	0.984
c_5	0.00609	0.00234	2.604	0.009***
c_6	-0.00039	0.00234	-0.165	0.869
c_7	0.00277	0.00234	1.188	0.235

Note:

1. the average daily return is 0.002357.

2. *, ** and *** shown in *p*-value column denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4-6 Regression analysis results of intraday effect on Bitcoin price return

parameter	coefficient	sdv	t-statistic	p- value
d_1	-0.00005	0.00032	-0.166	0.868
d_2	-0.00024	0.00032	-0.743	0.458
d_3	-0.00017	0.00032	-0.531	0.595
d_4	0.00031	0.00032	0.978	0.328
d_5	-0.00010	0.00032	-0.311	0.756
d_6	0.00074	0.00032	2.318	0.020**
d_7	-0.00017	0.00032	-0.543	0.587
d_8	0.00056	0.00032	1.750	0.080*
d_9	0.00005	0.00032	0.151	0.880
d_{10}	-0.00008	0.00032	-0.246	0.806
d_{11}	0.00105	0.00032	3.290	0.001***
d_{12}	0.00035	0.00032	1.099	0.272

Note: the average two-hour return is 0.000182

Table 4-7 Power ratio analysis results of weekday effect on Bitcoin price return

Weekday	2015	2016	2017	2018	2019	2020	2021
Monday	1.103	0.989	1.054	0.930	1.027	1.051	0.960
Tuesday	0.945	0.970	1.012	1.026	0.969	1.013	0.996
Wednesday	1.060	1.021	1.001	0.968	0.996	1.018	1.037
Thursday	0.966	1.013	1.020	0.981	0.933	0.975	0.986
Friday	1.101	1.015	0.917	1.087	1.111	1.002	1.023
Saturday	0.915	0.994	0.978	0.971	0.984	0.994	0.983
Sunday	0.938	1.007	1.027	1.043	0.990	0.963	1.006

Table 4-8 Power ratio analysis results of intraday effect on Bitcoin price return

O'clock	2015	2016	2017	2018	2019	2020	2021
0-2	1.002	1.001	1.032	1.006	1.010	1.000	1.001
2-4	1.004	1.003	1.011	0.995	1.009	1.008	1.001
4-6	1.010	1.007	1.000	0.989	0.999	0.996	0.999
6-8	1.004	0.997	1.010	0.985	0.996	0.998	1.000
8-10	1.000	0.994	0.998	1.008	0.989	1.001	1.000
10-12	1.016	1.004	1.022	0.990	0.997	0.994	1.000
12-14	1.026	1.007	1.000	0.984	1.003	0.987	1.001
14-16	1.039	1.003	1.001	1.020	1.001	0.994	0.999
16-18	0.979	1.004	0.998	0.992	1.001	0.999	1.000
18-20	1.005	0.997	1.020	0.999	1.012	0.983	1.000
20-22	1.002	1.011	0.996	0.991	1.002	0.982	0.999
22-24	0.994	1.001	1.000	0.997	1.002	1.012	0.999

Table 4-9 Partial rendering of database prepared for correlation analysis of tweets/Bitcoin price return

Time	N	$F_{\_AVG}$	$Q_{\_AVG}$	$RT_{\_AVG}$	$SC_{\_AVG}$	$SC{*}F_{\_AVG}$	$SC{*}Q_{\_AVG}$	$SC{*}RT_{\_AVG}$	$SC_{\_sdv}$	$SC{*}F_{\_sdv}$	$SC{*}Q_{\_sdv}$	$SC{*}RT_{\_sdv}$	R	Vol
10/1 00:00:00 ~ 10/1 00:09:59	696	114.101	2.856	24.249	0.238	75.559	2.161	16.528	0.423	1802.412	48.629	399.749	-0.0011	24.35
10/1 00:10:00 ~ 10/1 00:19:59	572	14.119	2.290	7.598	0.327	2.755	-0.218	5.772	0.421	67.767	5.265	138.504	0.0009	7.25
10/1 00:20:00 ~ 10/1 00:29:59	510	27.102	0.490	4.345	0.290	0.519	0.007	0.287	0.446	17.060	0.740	7.808	0.0014	9.54
: ; ;	: ; ;	: ; ;	: ; ;	: ; ;	: ; ;	: ; ;	: ; ;	: ; ;	: ; ;	: ; ;	: ; ;	: ; ;	: ; ;	: ; ;
10/30 23:30:00 ~ 10/30 23:39:59	595	23.345	0.440	4.575	0.320	-1.744	-0.074	-0.538	0.418	109.095	2.090	24.664	0.0013	21.79
10/30 23:40:00 ~ 10/30 23:49:59	652	13.505	3.104	4.952	0.362	4.357	2.088	2.491	0.442	93.690	53.052	59.046	-0.0001	0.68
10/30 23:50:00 ~ 10/30 23:59:59	571	30.420	0.681	5.356	0.283	1.267	0.014	0.402	0.421	25.588	0.189	8.131	0.0027	12.36

Table 4-10 Statistics of the variables of concern in this research

	mean	sdv	max	min	skewness	kurtosis
N	842.8474	320.1097	7273.0000	205.0000	4.7802	60.5246
$F_{\_avg}$	29.0728	79.0354	3390.2924	0.5402	22.6668	822.6587
$Q_{\_avg}$	1.9346	5.4897	196.9930	0.0108	15.7561	446.3714
$RT_{\_avg}$	7.6911	14.6907	608.9022	0.2165	18.8052	676.3288
$SC_{\_avg}$	0.2932	0.0603	0.5221	-0.1000	-0.1737	0.9939
$SC{*}F_{\_avg}$	5.0774	23.7335	863.8589	-643.5862	7.9312	548.6948
$SC{*}Q_{\_avg}$	0.6519	2.5134	54.7711	-20.5679	7.4251	104.6010
$SC{*}RT_{\_avg}$	2.0816	5.4672	66.3059	-117.0713	0.4705	77.2312
$SC_{\_sdv}$	0.4299	0.0235	0.5109	0.3495	-0.0529	-0.0573
$SC{*}F_{\_sdv}$	190.0332	589.4334	18381.7140	0.9875	16.0629	404.4814
$SC{*}Q_{\_sdv}$	21.6945	68.6920	1798.3971	0.0260	9.7965	166.0829
$SC{*}RT_{\_sdv}$	66.9483	136.7258	3147.6412	0.3464	6.7671	89.0919
$R_{\_AVG}$	0.0001	0.0029	0.0356	-0.0531	-0.5308	43.2446
R_sdv	0.0007	0.0006	0.0144	0.0001	8.3187	145.5817
Vol	11.6649	27.2885	827.8801	0.0028	16.7427	413.2104

Table 4-11 Correlation analysis results of Bitcoin return and number of tweets for different time lags

N	-30	-20	-10	0	10	20	30
coefficient	3.8E-07	5.4E-07	5.6E-07	4.1E-07	1.4E-07	6.2E-09	-5.2E-08
sdv	1.4E-07	1.4E-07	1.4E-07	1.4E-07	1.4E-07	1.4E-07	1.4E-07
t-statistic	2.722	3.898	4.077	2.982	0.992	0.045	-0.379
p- value	0.007	0.000	0.000	0.003	0.321	0.964	0.705
R^2	0.042	0.059	0.062	0.045	0.015	0.001	0.006

Table 4-12 Correlation analysis results of Bitcoin return and average sentiment compound score for different time lags

SC_{avg}	-30	-20	-10	0	10	20	30
coefficient	-7.3E-04	-1.9E-03	-1.6E-03	-1.1E-03	-6.9E-05	3.2E-04	2.3E-04
sdv	7.3E-04	7.3E-04	7.3E-04	7.3E-04	7.3E-04	7.3E-04	7.3E-04
t-statistic	-0.993	-2.647	-2.250	-1.468	-0.094	0.440	0.312
p- value	0.321	0.008	0.024	0.142	0.925	0.660	0.755
R^2	0.015	0.040	0.034	0.022	0.001	0.007	0.005

Table 4-13 Correlation analysis results of Bitcoin return and average sentiment compound score weighted by favorite counts for different time lags

$SC*F\_{avg}$	-30	-20	-10	0	10	20	30
coefficient	7.8E-07	-1.0E-06	1.4E-06	2.9E-07	1.3E-06	-6.0E-07	-4.5E-07
sdv	1.9E-06	1.9E-06	1.9E-06	1.9E-06	1.9E-06	1.9E-06	1.9E-06
t-statistic	0.419	-0.555	0.742	0.157	0.723	-0.320	-0.242
p- value	0.675	0.579	0.458	0.876	0.470	0.749	0.809
R^2	0.006	0.008	0.011	0.002	0.011	0.005	0.004

Table 4-14 Correlation analysis results of Bitcoin return and average sentiment compound score weighted by quote counts for different time lags

SC*Q_avg	-30	-20	-10	0	10	20	30
coefficient	4.5E-07	-1.5E-05	-1.3E-05	1.1E-05	4.3E-06	-4.0E-06	-6.2E-06
sdv	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05
t-statistic	0.025	-0.846	-0.711	0.646	0.247	-0.227	-0.352
p- value	0.980	0.398	0.477	0.518	0.805	0.821	0.725
R^2	0.000	0.013	0.011	0.010	0.004	0.003	0.005

Table 4-15 Correlation analysis results of Bitcoin return and average sentiment compound score weighted by retweet counts for different time lags

SC*RT_avg	-30	-20	-10	0	10	20	30
coefficient	-7.3E-04	-1.5E-05	1.3E-06	3.2E-06	1.3E-05	-1.4E-05	1.0E-06
sdv	7.3E-04	8.1E-06	8.1E-06	8.1E-06	8.1E-06	8.1E-06	8.1E-06
t-statistic	-0.993	-1.806	0.163	0.394	1.620	-1.717	0.124
p- value	0.321	0.071	0.871	0.694	0.105	0.086	0.901
R^2	0.019	0.028	0.002	0.006	0.025	0.026	0.002

Table 4-16 Correlation analysis results of Bitcoin return and standard deviation of sentiment compound score for different time lags

SC_sdv	-30	-20	-10	0	10	20	30
coefficient	1.0E-05	1.2E-03	-4.1E-03	1.6E-03	4.8E-03	-3.1E-03	4.1E-04
sdv	8.1E-06	1.9E-03	1.9E-03	1.9E-03	1.9E-03	1.9E-03	1.9E-03
t-statistic	1.249	0.615	-2.122	0.815	2.515	-1.642	0.215
p- value	0.212	0.538	0.034	0.415	0.012	0.101	0.830
R^2	0.004	0.009	0.032	0.012	0.038	0.025	0.003

Table 4-17 Correlation analysis results of Bitcoin return and standard deviation of sentiment compound score weighted by favorite counts for different time lags

SC*F_sdv	-30	-20	-10	0	10	20	30
coefficient	-5.9E-08	-5.2E-09	9.0E-08	5.3E-08	1.4E-07	8.1E-09	-4.7E-08
sdv	7.5E-08	7.5E-08	7.5E-08	7.5E-08	7.5E-08	7.5E-08	7.5E-08
t-statistic	-0.779	-0.069	1.203	0.710	1.854	0.108	-0.627
p- value	0.436	0.945	0.229	0.478	0.064	0.914	0.530
R^2	0.012	0.001	0.018	0.011	0.028	0.002	0.010

Table 4-18 Correlation analysis results of Bitcoin return and standard deviation of sentiment compound score weighted by quote counts for different time lags

SC*Q_sdv	-30	-20	-10	0	10	20	30
coefficient	-3.3E-07	-3.0E-07	3.9E-08	5.0E-07	3.3E-07	-1.6E-07	-6.5E-07
sdv	6.4E-07	6.4E-07	6.4E-07	6.4E-07	6.4E-07	6.4E-07	6.4E-07
t-statistic	-0.515	-0.467	0.060	0.780	0.516	-0.247	-1.015
p- value	0.607	0.640	0.952	0.436	0.606	0.805	0.310
R^2	0.008	0.007	0.001	0.012	0.008	0.004	0.015

Table 4-19 Correlation analysis results of Bitcoin return and standard deviation of sentiment compound score weighted by retweet counts for different time lags

SC*RT_sdv	-30	-20	-10	0	10	20	30
coefficient	1.3E-07	-4.9E-07	3.4E-07	4.1E-07	7.2E-07	-4.7E-07	-1.9E-07
sdv	3.2E-07	3.2E-07	3.2E-07	3.2E-07	3.2E-07	3.2E-07	3.2E-07
t-statistic	0.397	-1.524	1.044	1.273	2.240	-1.450	-0.592
p- value	0.691	0.128	0.296	0.203	0.025	0.147	0.554
R^2	0.006	0.023	0.016	0.019	0.034	0.022	0.009

Table 4-20 Correlation analysis results of Bitcoin return and trade volume for different time lags

Vol.	-30	-20	-10	0	10	20	30
coefficient	1.5E-06	5.8E-06	-1.2E-06	8.2E-06	1.4E-05	2.0E-06	-1.7E-06
sdv	1.6E-06	1.6E-06	1.6E-06	1.6E-06	1.6E-06	1.6E-06	1.6E-06
t-statistic	0.930	3.554	-0.716	5.067	8.397	1.209	-1.054
p- value	0.352	0.000	0.474	0.000	0.000	0.227	0.292
R^2	0.014	0.054	0.011	0.077	0.127	0.018	0.016

Table 4-21 Correlation analysis results of volatility in Bitcoin return and number of tweets for different time lags

N	-30	-20	-10	0	10	20	30
coeff.	3E-07	3.26E-07	3.6E-07	4.7E-07	4.57E-07	4.2E-07	3.81E-07
sdv	2.66E-08	2.65E-08	2.64E-08	2.6E-08	2.6E-08	2.62E-08	2.63E-08
t-statistic	11.28913	12.29615	13.62053	18.273	17.5457	16.03587	14.46637
p- value	3.81E-29	3.55E-34	2.16E-41	0.000	1.29E-66	3E-56	2.39E-46
R^2	0.169796	0.184427	0.203484	0.269	0.258617	0.237702	0.215597

Table 4-22 Correlation analysis results of volatility in Bitcoin return and average sentiment compound score for different time lags

SC_avg	-30	-20	-10	0	10	20	30
coeff.	-6.73E-04	-8.43E-04	-1.06E-03	-1.44E-03	-1.36E-03	-1.27E-03	-1.21E-03
sdv	1.43E-04	1.42E-04	1.42E-04	1.41E-04	1.42E-04	1.42E-04	1.42E-04
t-statistic	-4.71E+00	-5.92E+00	-7.42E+00	-1.02E+01	-9.64E+00	-8.99E+00	-8.53E+00
p- value	2.52E-06	3.49E-09	1.36E-13	5.75E-24	8.61E-22	3.68E-19	1.99E-17
R	7.17E-02	9.00E-02	1.13E-01	1.53E-01	1.46E-01	1.36E-01	1.29E-01

Table 4-23 Correlation analysis results of volatility in Bitcoin return and average sentiment compound score weighted by favorite counts for different time lags

$SC{*}F_{\_avg}$	-30	-20	-10	0	10	20	30
coeff.	2.55E-07	3.80E-07	2.86E-07	9.54E-08	-3.63E-08	1.67E-07	3.36E-08
sdv	3.64E-07	3.64E-07	3.64E-07	3.64E-07	3.64E-07	3.64E-07	3.64E-07
t-statistic	7.02E-01	1.05E+00	7.88E-01	2.62E-01	-9.97E-02	4.60E-01	9.22E-02
p- value	4.83E-01	2.96E-01	4.31E-01	7.93E-01	9.21E-01	6.46E-01	9.27E-01
R	1.07E-02	1.60E-02	1.20E-02	4.00E-03	1.52E-03	7.02E-03	1.41E-03

Table 4-24 Correlation analysis results of volatility in Bitcoin return and average sentiment compound score weighted by quote counts for different time lags

$SC{*}Q_{\_avg}$	-30	-20	-10	0	10	20	30
coeff.	3.32E-06	-1.16E-06	1.36E-06	1.51E-06	2.17E-06	-4.07E-08	-4.18E-06
sdv	3.43E-06	3.43E-06	3.43E-06	3.43E-06	3.43E-06	3.43E-06	3.43E-06
t-statistic	9.66E-01	-3.38E-01	3.95E-01	4.41E-01	6.32E-01	-1.18E-02	-1.22E+00
p- value	3.34E-01	7.35E-01	6.93E-01	6.59E-01	5.27E-01	9.91E-01	2.24E-01
R	1.47E-02	5.17E-03	6.02E-03	6.72E-03	9.65E-03	1.81E-04	1.86E-02

Table 4-25 Correlation analysis results of volatility in Bitcoin return and average sentiment compound score weighted by retweet counts for different time lags

SC*RT_avg	-30	-20	-10	0	10	20	30
coeff.	3.57E-06	1.79E-06	5.00E-06	5.05E-07	8.19E-07	1.38E-06	-1.09E-06
sdv	1.58E-06	1.58E-06	1.58E-06	1.58E-06	1.58E-06	1.58E-06	1.58E-06
t-statistic	2.26E+00	1.14E+00	3.17E+00	3.20E-01	5.19E-01	8.71E-01	-6.87E-01
p- value	2.38E-02	2.56E-01	1.55E-03	7.49E-01	6.04E-01	3.84E-01	4.92E-01
R	3.45E-02	1.73E-02	4.83E-02	4.88E-03	7.91E-03	1.33E-02	1.05E-02

Table 4-26 Correlation analysis results of volatility in Bitcoin return and standard deviation of sentiment compound score for different time lags

$SC_{\_sdv}$	-30	-20	-10	0	10	20	30
coeff.	0.000228	-8.1E-05	-0.00085	-1.7E-05	5.28E-05	0.000321	0.000101
sdv	0.000374	0.000369	0.000366	3.8E-04	0.000368	0.000369	0.000368
t-statistic	0.609087	-0.22044	-2.31803	-0.045	0.143578	0.870113	0.274443
p- value	0.542499	0.82554	0.020494	0.964	0.88584	0.384287	0.783758
R^2	0.009296	0.003364	0.035348	0.001	0.002191	0.013277	0.004189

Table 4-27 Correlation analysis results of volatility in Bitcoin return and standard deviation of sentiment compound score weighted by favorite counts for different time lags

$SC*F_{\_sdv}$	-30	-20	-10	0	10	20	30
coeff.	2.54E-08	1.85E-08	1.91E-08	1.88E-08	2.26E-08	2.71E-09	5.66E-09
sdv	1.46E-08	1.46E-08	1.46E-08	1.46E-08	1.47E-08	1.47E-08	1.47E-08
t-statistic	1.732957	1.260428	1.307863	1.285924	1.541932	0.18522	0.386321
p- value	0.083175	0.207584	0.19099	0.198539	0.123164	0.853065	0.699278
R^2	0.02644	0.019231	0.019952	0.019616	0.023521	0.002827	0.005896

Table 4-28 Correlation analysis results of volatility in Bitcoin return and standard deviation of sentiment compound score weighted by quote counts for different time lags

$SC*Q_{\_sdv}$	-30	-20	-10	0	10	20	30
coeff.	2.82E-07	1.25E-08	1.24E-07	2.73E-07	3.35E-07	8.71E-08	-3.4E-09
sdv	1.26E-07	1.26E-07	1.26E-07	1.26E-07	1.26E-07	1.26E-07	1.26E-07
t-statistic	2.245784	0.099613	0.987122	2.170118	2.667332	0.693516	-0.02745
p- value	0.024769	0.920656	0.323639	0.030053	0.007674	0.488024	0.978105
R^2	0.034256	0.00152	0.015061	0.033091	0.040666	0.010583	0.000419

Table 4-29 Correlation analysis results of volatility in Bitcoin return and standard deviation of sentiment compound score weighted by retweet counts for different time lags

SC*RT_sdv	-30	-20	-10	0	10	20	30
coeff.	2.27E-07	8.27E-08	2.56E-07	1.69E-07	1.69E-07	7.63E-08	1.42E-08
sdv	6.3E-08	6.31E-08	6.3E-08	6.31E-08	6.31E-08	6.32E-08	6.32E-08
t-statistic	3.598752	1.310722	4.060769	2.684475	2.684475	1.208284	0.225012
p- value	0.000323	0.190022	4.98E-05	0.007292	0.007292	0.227005	0.821981
R^2	0.054843	0.019998	0.061844	0.040927	0.040927	0.018436	0.003434

Table 4-30 Correlation analysis results of volatility in Bitcoin return and trade volume for different time lags

Vol.	-30	-20	-10	0	10	20	30
coeff.	3.05E-06	4.47E-06	5.5E-06	1.46E-05	5.33E-06	3.78E-06	2.43E-06
sdv	3.13E-07	3.09E-07	3.05E-07	2.25E-07	3.06E-07	3.11E-07	3.14E-07
t-statistic	9.752625	14.45515	18.01721	64.98389	17.4416	12.14575	7.739664
p- value	3.05E-22	2.79E-46	5.1E-70	0	7.08E-66	2.12E-33	1.23E-14
R^2	0.147225	0.215414	0.265084	0.704067	0.257184	0.182246	0.117309

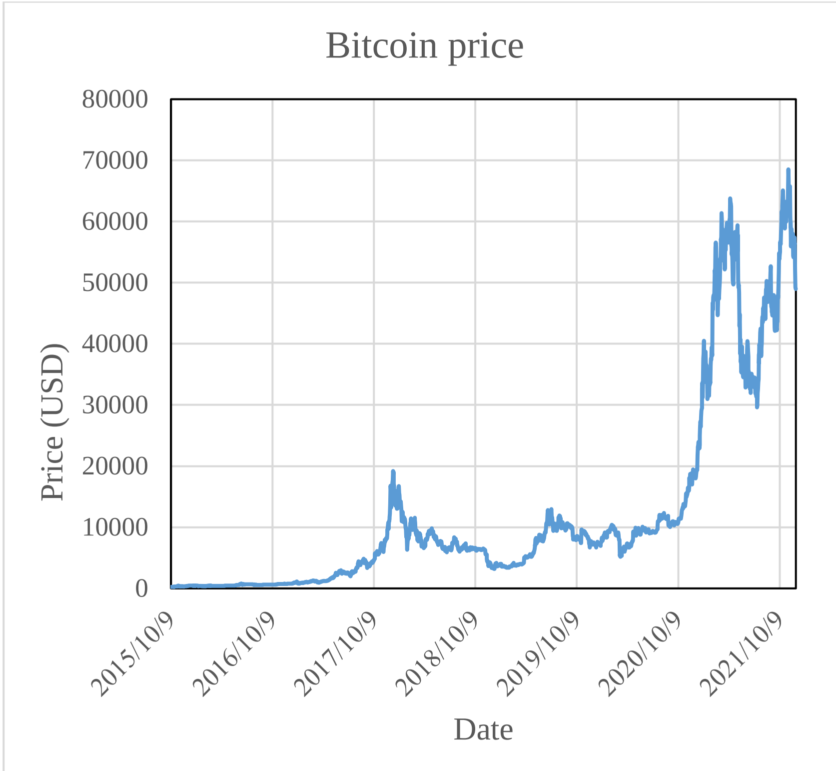

Figure 4-1 Time series of Bitcoin price between 2015/10/9 and 2021/10/9

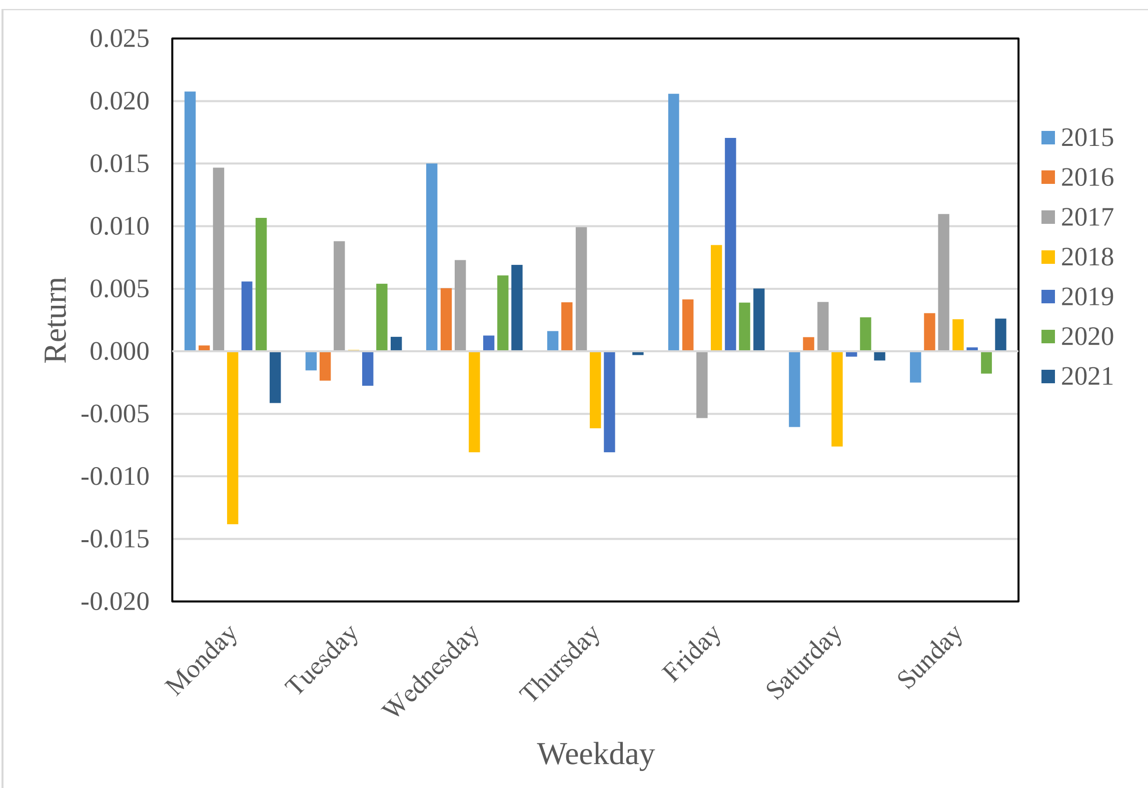

Figure 4-2 Daily returns of Bitcoin price (Dataset_1)

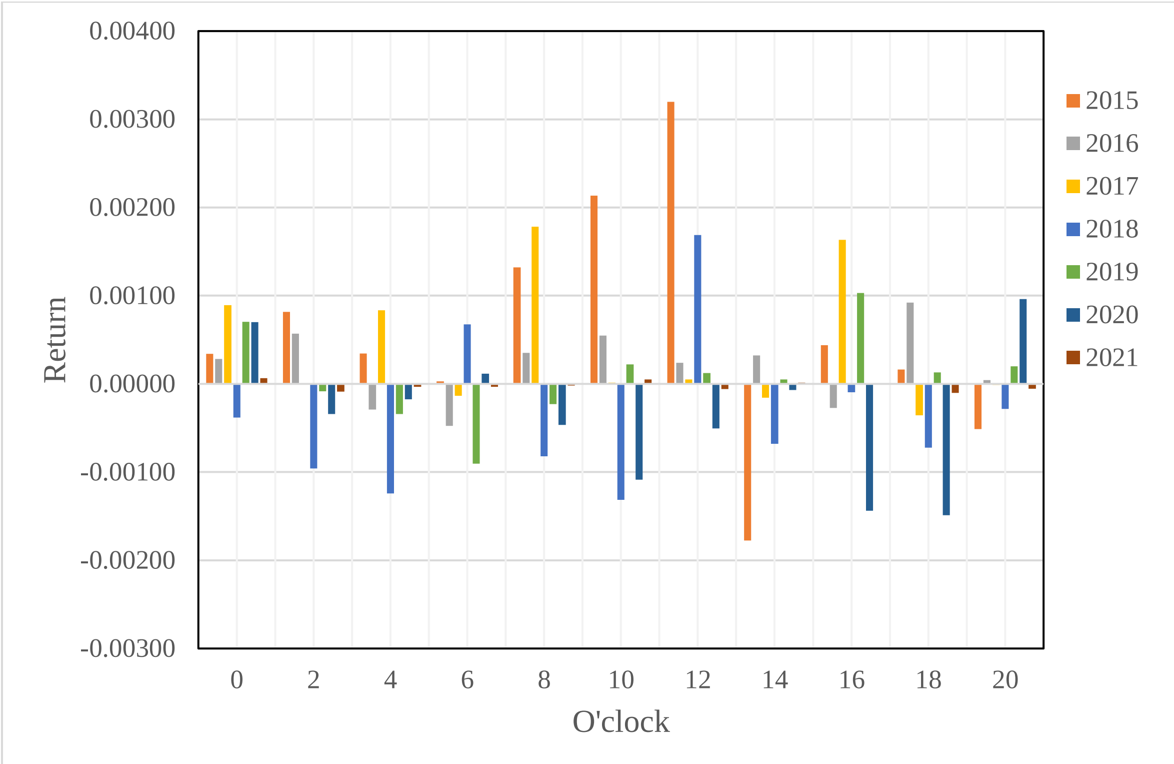

Figure 4-3 Intraday returns of Bitcoin price (Dataset_2)

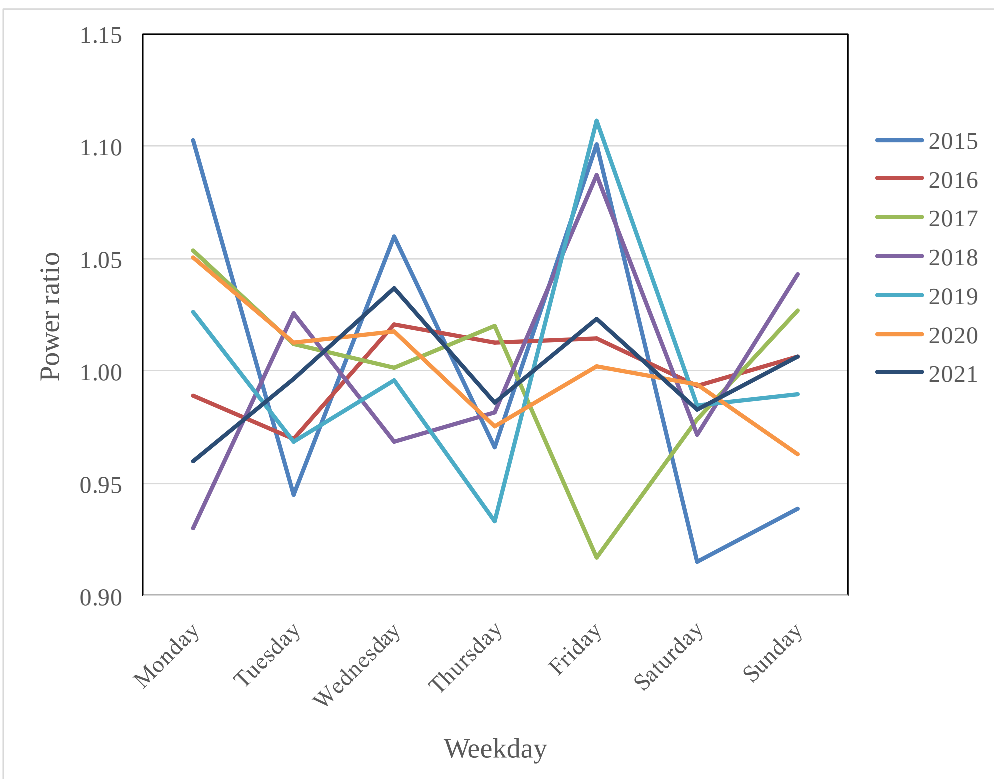

Figure 4-4 Weekday power ratios of Bitcoin price return over 2015 to 2021

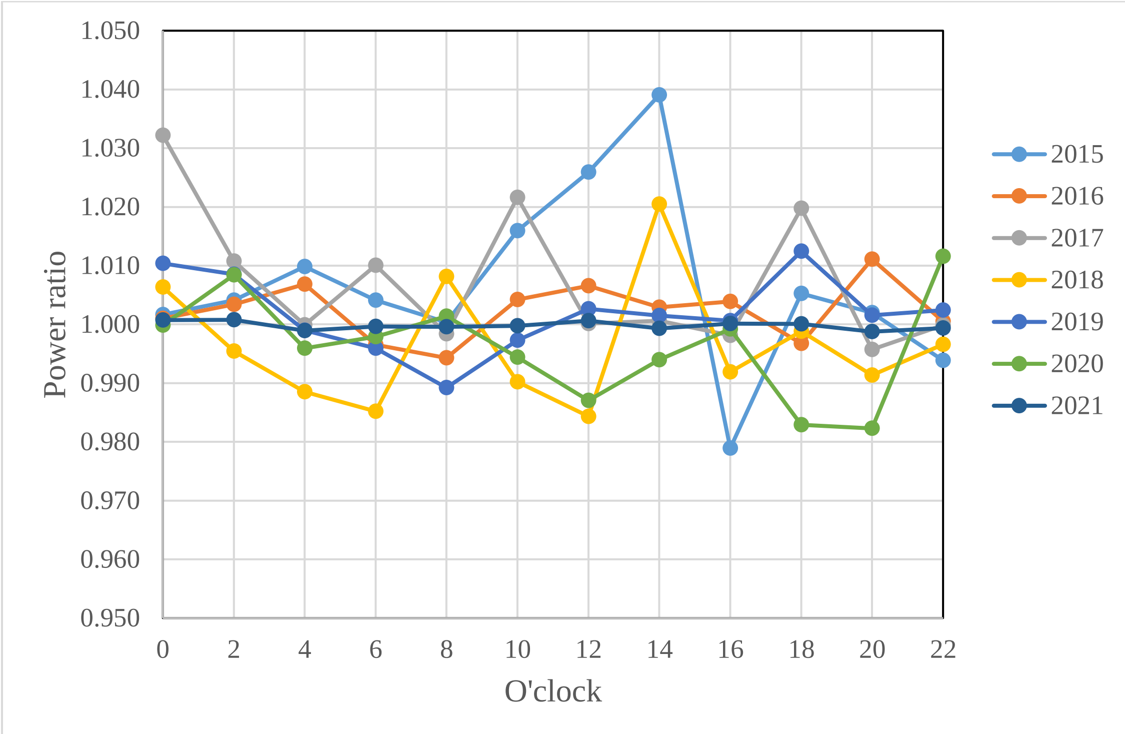

Figure 4-5 Intraday power ratios of Bitcoin price return over 2015 to 2021

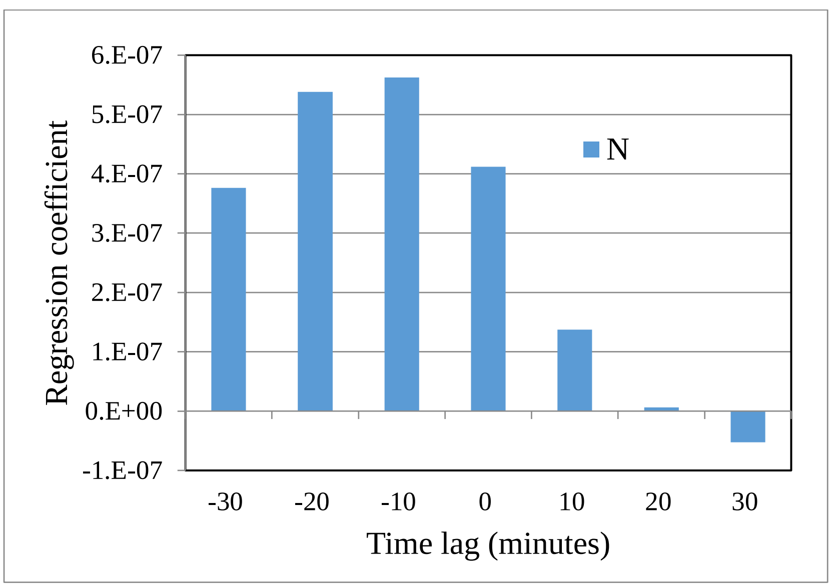

Figure 4-6 Regression coefficients of Bitcoin return and number of tweets for different time lags

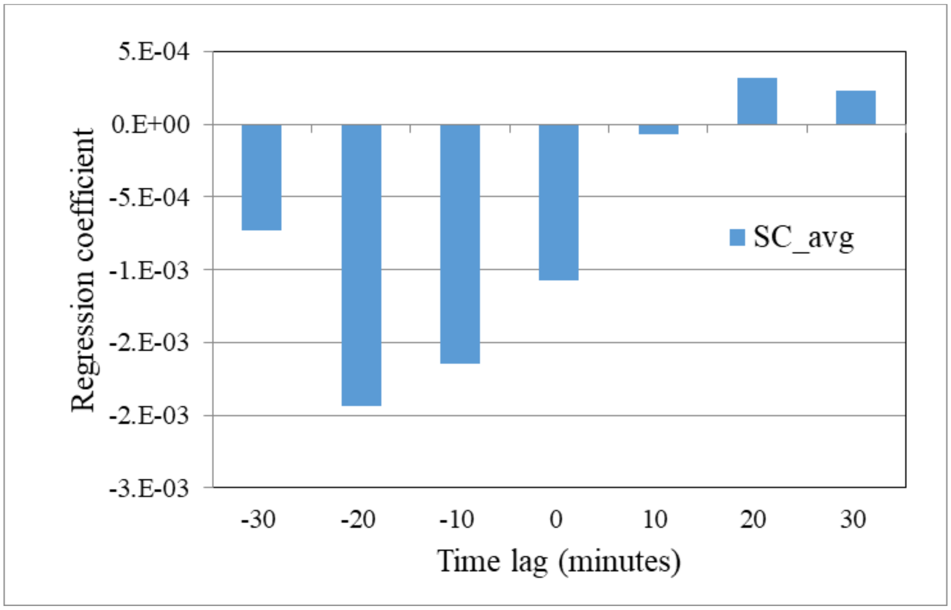

Figure 4-7 Regression coefficients of Bitcoin return and average sentiment compound score for different time lags

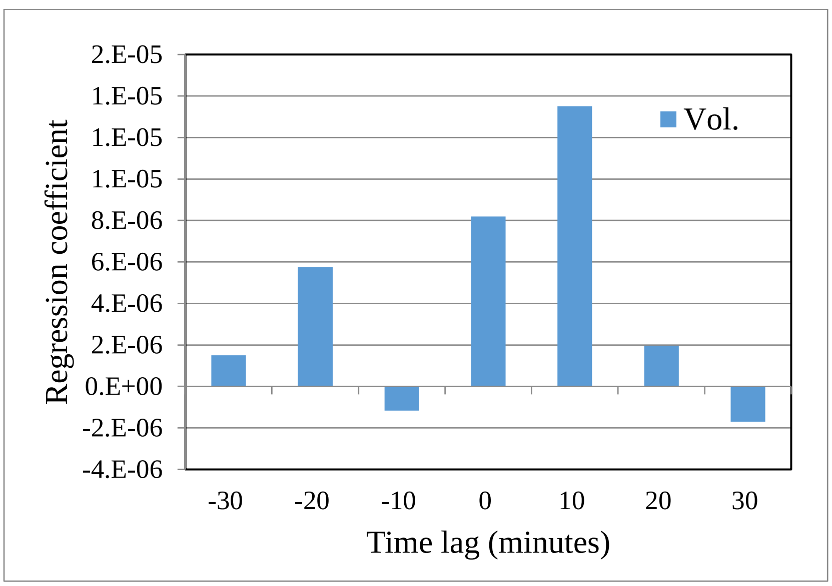

Figure 4-8 Regression coefficients of Bitcoin return and trading volume for different time lags

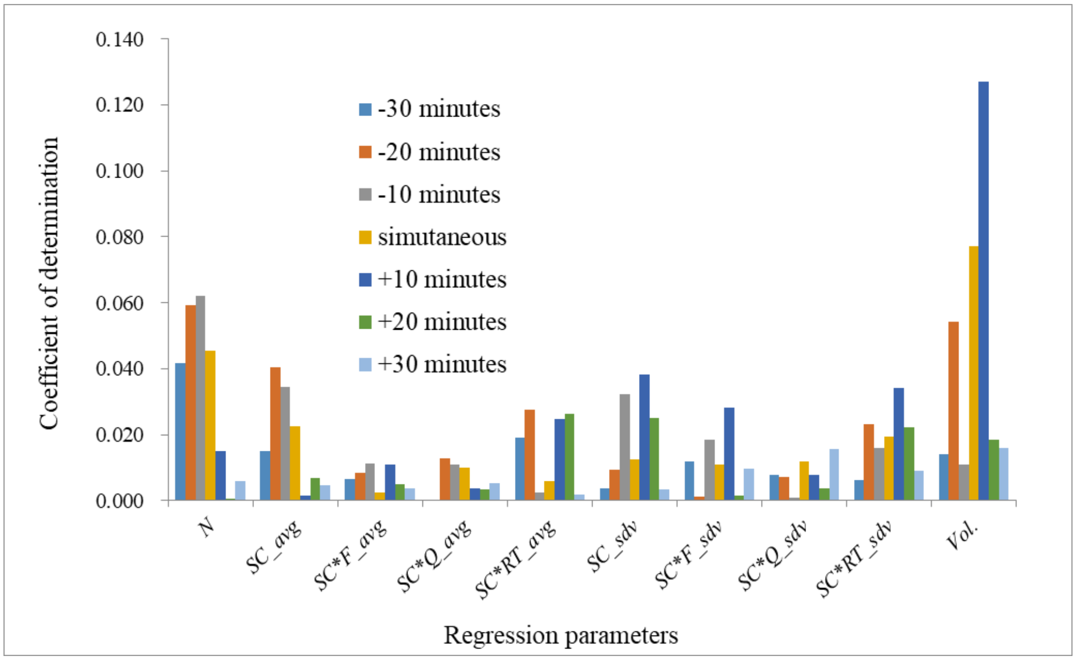

Figure 4-9 Coefficients of determination for correlation analysis of Bitcoin return and variant parameters for different time lags

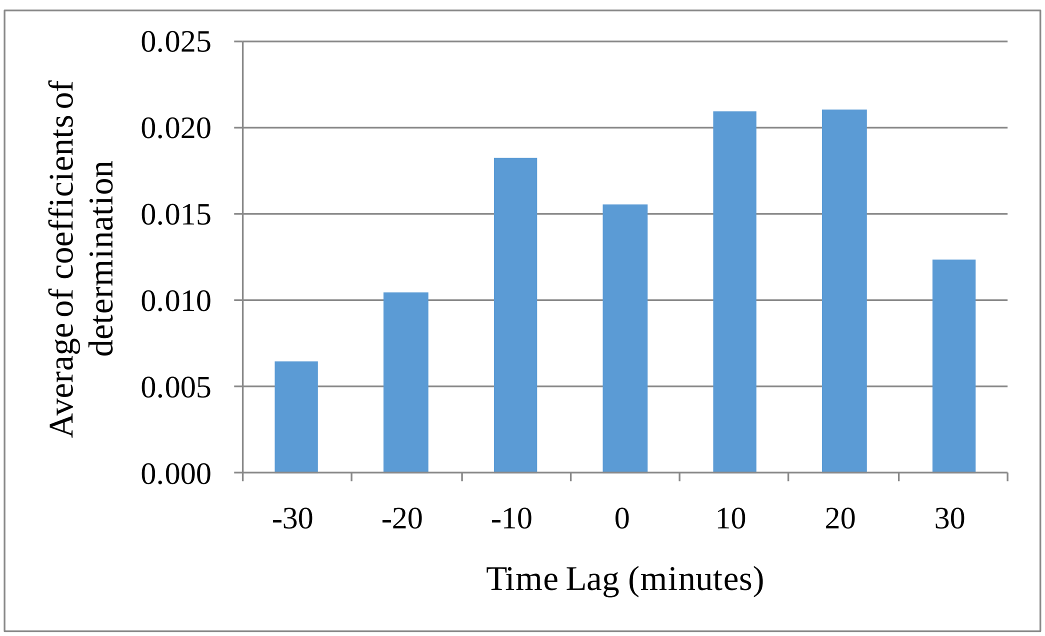

Figure 4-10 Average coefficient of determination for 9 tweet parameters for different time lags, for regression analysis of Bitcoin return

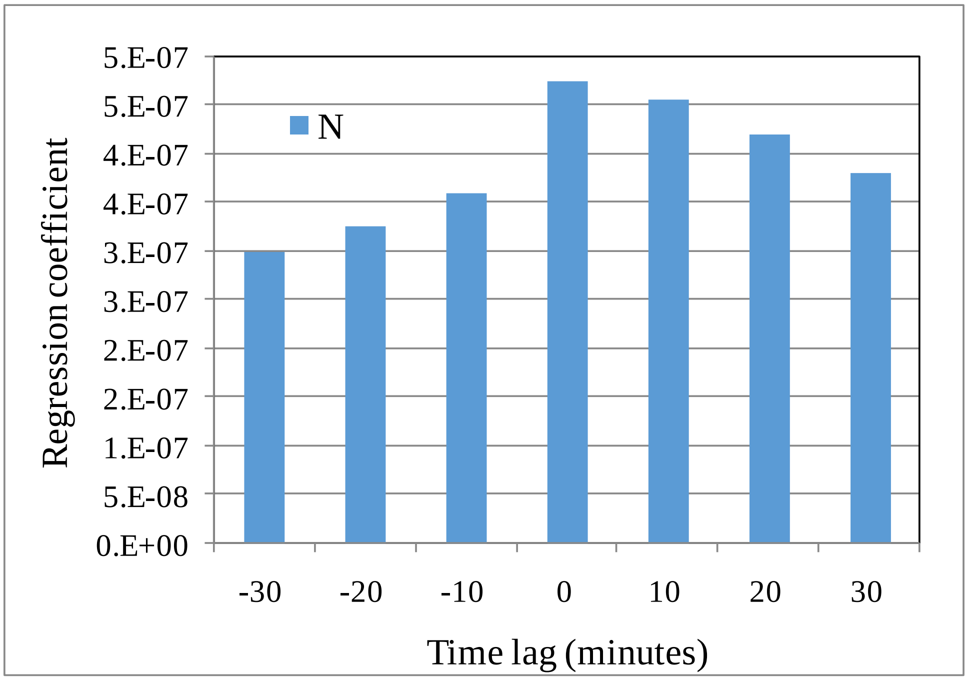

Figure 4-11 Regression coefficients of volatility in Bitcoin return and number of tweets for different time lags

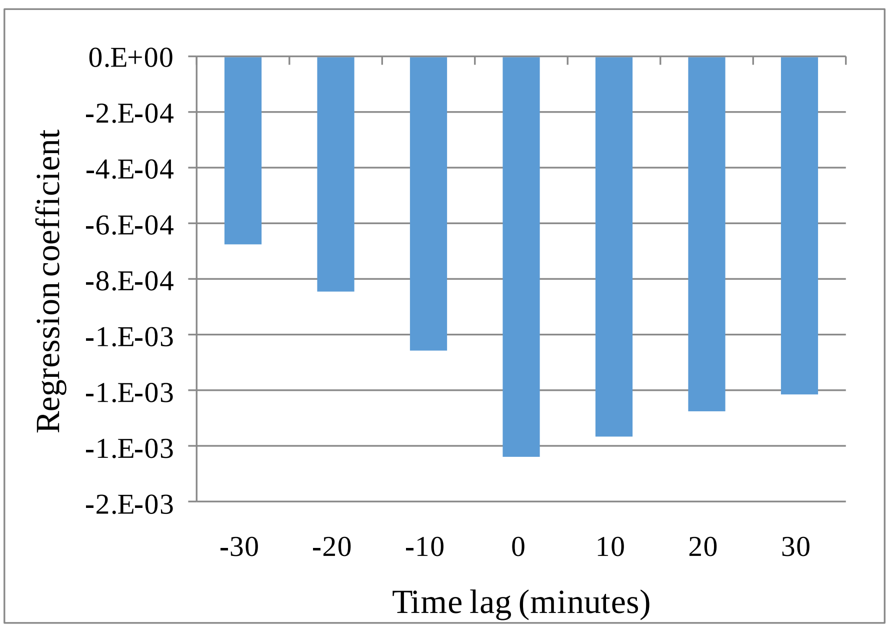

Figure 4-12 Regression coefficients of volatility in Bitcoin return and average sentiment compound score for different time lags

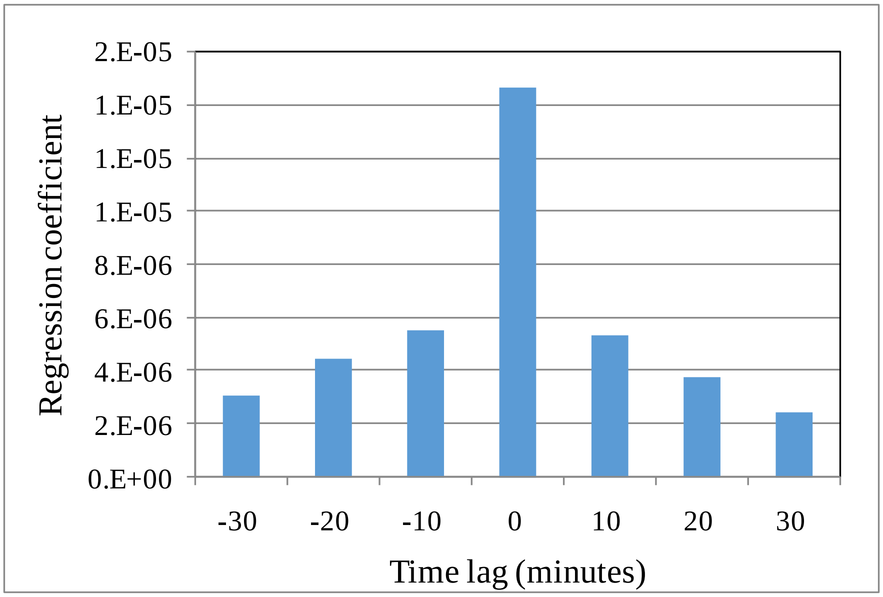

Figure 4-13 Regression coefficients of volatility in Bitcoin return and trading volume for different time lags

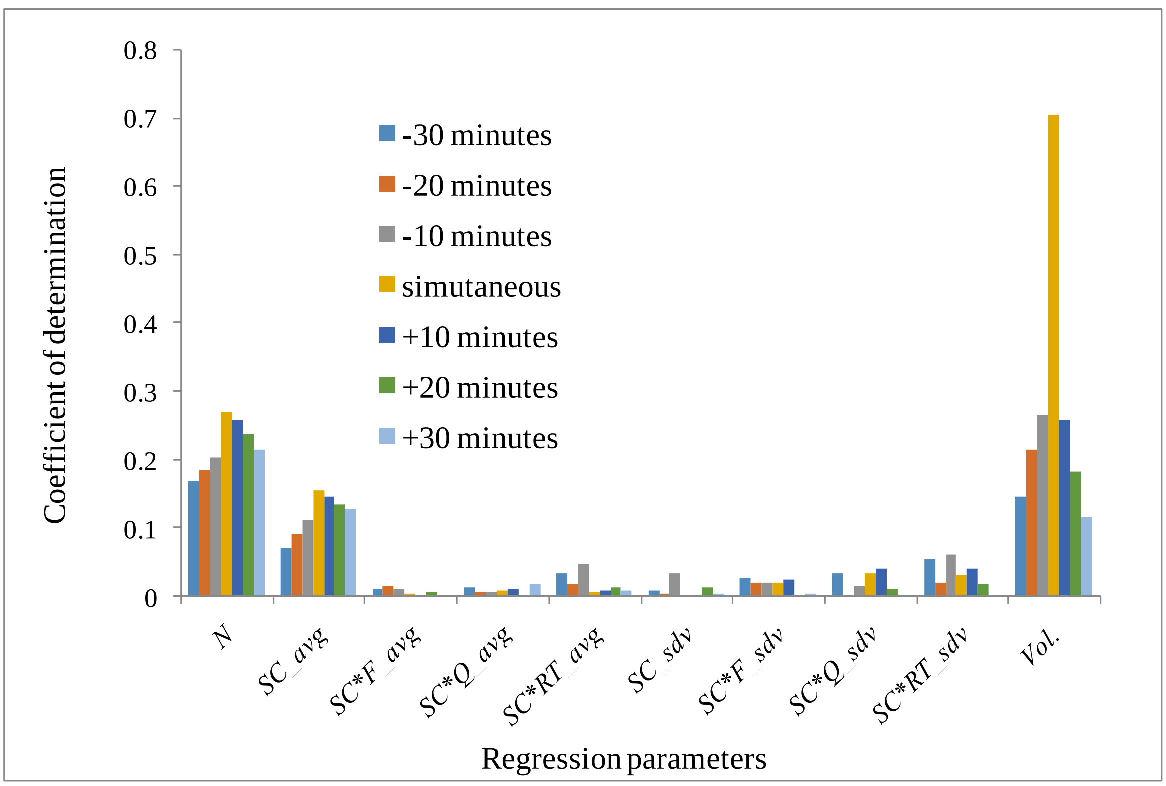

Figure 4-14 Coefficients of determination for correlation analysis of volatility in Bitcoin return and variant parameters for different time lags

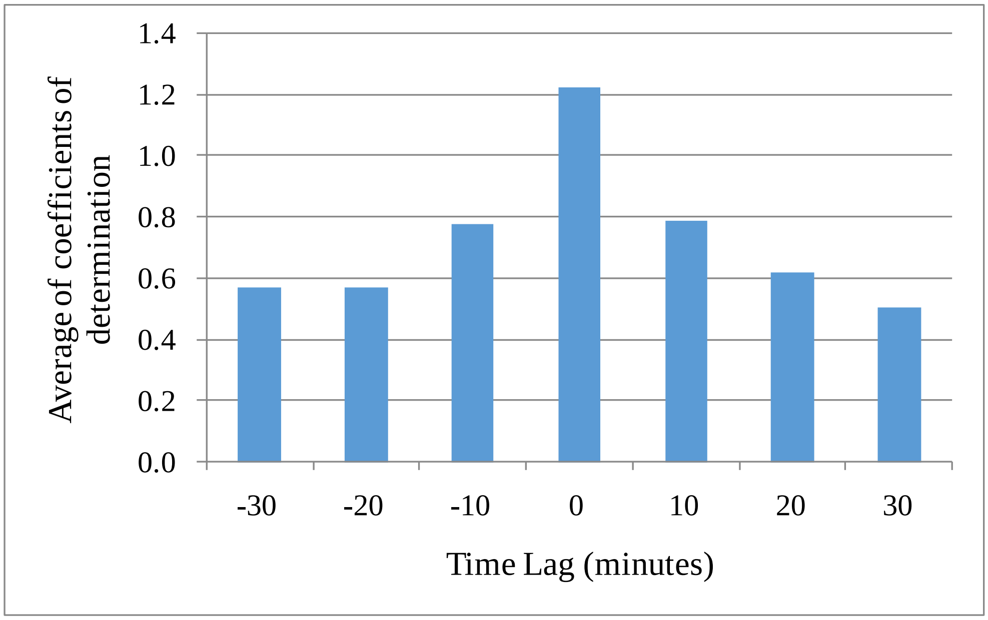

Figure 4-15 Average coefficient of determination for 9 tweet parameters for different time lags, for regression analysis of volatility in Bitcoin return

Chapter 5

Discussions and Conclusions

Investors aim to profit from their decisions over investment assets, and Bitcoin is one of the most popular assets that have emerged in the last decade. This research examines if Bitcoin's price satisfies the efficient market hypothesis (EMH) and specifically examines if the calendar effect is significant for it. The calendar effect can be used to obtain abnormal profits, and any sentiment effect in Bitcoin can also be integrated into profit models.

The day-of-the-week and intraday effects are studied herein on daily and hourly Bitcoin price data spanning between October 2015 and November 2021. Regression analysis with dummy variables and power ratio analysis are then conducted. Regression analysis results reveal that, except for Friday, no anomaly is detected on the other days of the week. For the intraday effect, some positive intraday anomalies are found at 10:00-12:00, 14:00-16:00, and 20:00-22:00.These analysis results give credence that Bitcoin returns do not conform to the efficient market with respect to day-of-the-week and intraday anomalies. However, based on the inferred statistics of power ratio analysis, the anomaly between day-of-the-week and intraday is not significant. More contradictory, power ratio analysis shows positive returns in the early morning, while regression analysis gives the opposite results. It is interesting to find that varied conclusions arise based on the same dataset, but different research methods. It explains why recent research results about calendar anomalies in Bitcoin returns refute each other. The significance of the day-of-the-week and intraday effect of Bitcoin is still not clear.

As Bitcoin is a purely belief-driven asset and has no concrete fundamental value, optimistic sentiment over it can lead to an increase in its price, while the emergence of negative sentiment could erode the beliefs among Bitcoin investors that then lead to a situation in which disagreement predicts low returns. Therefore, it is of interest to study how sentiment-related information revealed in Twitter affects the return and volatility of Bitcoin.

This study of the sentiment effect is based on minute-level Bitcoin price data and Bitcoin-related tweets posted on Twitter in October 2021. In total, more than 4 million tweets are crawled and processed for sentiment compound scores, using the VADAR-based sentiment analysis package embedded in Python. The affiliated information, including favorites, quotes, and number of retweets, are also crawled for weighing their sentiment scores. The analysis results show that the information retrieved from Twitter, including number of tweets posted during the time period, unweighted sentiment compound score, sentiment compound score weighted by retweets, and trade volume, correlate more with Bitcoin returns. The number of tweets and trade volume positively correlate, while the unweighted and retweet weighted compound scores generally negatively correlate with Bitcoin returns. For the correlation with volatility in Bitcoin, a similar conclusion is reached except for the weighted compound scores is positively correlated. It is found that Twitter information correlates more with volatility than with Bitcoin returns, and the correlation is mostly with volatility simultaneously, but less correlated as time elapses.

A common proposition is that tweets about Bitcoins are representative for revealing the sentiment of Bitcoin investors. However, the analysis results indicate that the correlated effect of sentiment parameters, such as weighted and unweighted compound scores on Bitcoin price, is not consistent over the time horizon. This analysis provides a detailed investigation and discovers that when the time difference exceeds more than 20 minutes, no matter for leading or lagging, the correlation between Twitter information and Bitcoin price is less significant. Tweets' sentiment is a positive leading index, but it turns negative for a simultaneous or lagging index. This finding conforms to the fact that recent research results about the relationship between investor sentiment and Bitcoin contradict each other.

Considering their characteristics of real-time, representativeness, and quantity, the abundant messages on social media may provide an indicator for investor sentiment. For example, over 4 million posts are crawled from Twitter for sentiment analysis on Bitcoin. Some crawled tweets are accompanied by affiliated information, and the sentiment scores assessed by VADER analysis are extracted in the Appendix for reference. The main concern of this thesis is in the selection of crawling techniques and the sentiment analysis tool. Using different keywords on different social media platforms may scratch different posts, making it a potential issue for follow-up analysis. The same can be said towards the different sentiment analysis tools that may output different sentiment scores. The question of how to capture objective sentiment data from social media would be an interesting topic for further study.

One limitation of this research is that the calendar effect and sentiment analysis results are based on the data of the selected time period, and the results may not be appropriate for extrapolating to other durations. Another limitation of this research is that the p-values are not significant, and the R^2values are too insignificant for the single variable linear regression models examined in sentiment analysis. Some recent studies (e.g., Kristoufek 2013; Stavroyiannis et al. 2019) proposed that the relationship between cryptocurrency prices and investor sentiment is not linear. For example, Li et al. (2019) applied the Extreme Gradient Boosting Regression Tree Model to investigate investors' sentiment effect on cryptocurrency's price fluctuation. They reported that general sentiment is a powerful indicator that can better predict cryptocurrency price movements. This makes advanced models developed by machine learning a relevant approach to further study the lead-lag interactions between sentiment variables and Bitcoin price action.

Appendix A

Extracted Tweets Information and Sentiment Scores

Time	F	Q	RT	Pos	Neg	Neu	Comp	Text	Preprocessed text
2021-10-01 00:00:00				.0	.0	.0	.0	Top50 #Cryptocurrency IN/OUT update in last 12 hours (#crypto #bitcoin #altcoin): IN: $NEO, $KLAY OUT: $LEO, $UST https://t.co/u2nM5wvf3K	top50 inout update last 12 hours neo klay leo ust
2021-10-01 00:00:00				.326	.093	.581	.6124	What do you think about this cool #Bitcoin Warrior #NFT? You can get one of these for FREE through our #giveaway promo which is coming soon.. Follow us so you don't miss out!	think cool warrior get one free promo coming soon follow us nt miss
2021-10-01 00:00:00				.0	.0	.0	.0	#Cryptocurrencies Current Prices: #Bitcoin $ 43824.44 € 37859.21 #Ethereum $ 3001.14 € 2594.41 #Cardano $ 2.1155 € 1.8278 #blockchain #BTC #ETH #ADA #tokens #smartcontracts	current prices 4382444 € 3785921 300114 € 259441 21155 € 18278
2021-10-01 00:00:01				.424	.374	.202	.128	If you hate freedom... Don't buy #Bitcoin	hate freedom nt buy
2021-10-01 00:00:01				.0	.0	.0	.0	What price did you first buy #Bitcoin at?	price first buy
2021-10-01 00:00:01				.343	.084	.572	.8885	Bitcoin: $43829.34 💚 +72.47 last 1 Hour (+0.17%) 💔 -174.24 last 5 Hours (-0.4%) +2289.91 last 24 Hours (+5.51%) #BitcoinPriceUpdates #Bitcoin #HourlyCrypto	bitcoin 4382934 7247 last 1 hour 017 💔 17424 last 5 hours 04 💚 228991 last 24 hours 551 powered api

								#CryptoUpdates #Crypto Powered By @CryptoCompare API	
2021-10-01 00:00:01				.056	.183	.761	.5423	Fed's Powell has no intent to ban or crypto Powell testified before the House Financial Services Committee on Thursday on matters related to the economy and the Covid-19 pandemic. https://t.co/MP1F6XG6Xy	fed powell intent ban crypto powell testified house financial services committee thursday matters related economy covid19 pandemic
2021-10-01 00:00:01				.084	.0	.916	.296	"The average total coins generated across the network per day stays the same. Faster machines just get a larger share than slower machines. If everyone bought faster machines, they wouldn’t get more coins than before." ~ Satoshi Nakamoto on #Bitcoin	average total coins generated across network per day stays faster machines get larger share slower machines everyone bought faster machines ’ get coins satoshi nakamoto
2021-10-01 00:00:02				.457	.0	.543	.7506	6:00 PM >> $BTC Price: $43771 >> Hourly Sentiment: -67 | H.Change: 40 #bitcoin #hourlysentiment #bitcoinprice -- https://t.co/LPYvHCsPig	600 pm gt gt btc price 43771 gt gt hourly sentiment 67 hchange 40
2021-10-01 00:00:02				.0	.0	.0	.0	#Bitcoin is currently $43,830.4548	currently 438304548
2021-10-01 00:00:03				.0	.0	.0	.0	Current Bitcoin Price: USD $43,830.45 GBP £32,514.08 Euro €37,845.31 #bitcoin #btc $btc #btcusd #btcgbp #btceur #crypto #cryptocurrency	current bitcoin price usd 4383045 gbp £3251408 euro €3784531 btc
2021-10-01 00:00:03	8			.0	.0	.0	.0	IT'S OFFICIALLY Q4!!! #bitcoin	officially q4
2021-10-01 00:00:03				.0	.0	.0	.0	Buy Bitcoin $BTC @ 43824.1	buy bitcoin btc 438241
2021-10-01 00:00:03				.174	.0	.826	.2263	One Bitcoin now worth $43830.455. Market Cap $825.369 Billion. Based on #coindesk BPI #bitcoin	one bitcoin worth 43830455 market cap 825369 billion based bpi
2021-10-01 00:00:04				.136	.15	.714	.0516	One Bitcoin now worth $43834.43@bitstamp. High $44117.740. Low $41427.870. Market Cap $825.444 Billion #bitcoin	one bitcoin worth 4383443 high 44117740 low 41427870 market cap 825444 billion
2021-10-01 00:00:04				.0	.0	.0	.0	5 #cryptocurrency exchanges that the UK citizens can look to explore while considering #crypto	5 exchanges uk citizens look explore considering investments

								investments. #FCA #Bitcoin https://t.co/hXtWXk2eM8	
2021-10-01 00:00:05				.0	.0	.0	.0	The price of #Bitcoin is currently $43,858.37 #Crypto $BTC #BTC https://t.co/StFdwaoRx2	price currently 4385837 btc
2021-10-01 00:00:05	58		3	.0	.565	.435	.8442	Bitcoin Fear and Greed Index is 27 — Fear Current price: $43,820 https://t.co/jA72P4jZ79	bitcoin fear greed index 27 — fear current price 43820
2021-10-01 00:00:05				.234	.0	.766	.7003	LONG SIGNAL #BTC LONG BINANCE:BTCUSDTPERP, @ 43796.57 at 2021-10-01T00:00:00Z (Long-term signal appearing on the SwingSwiss daily chart at bar close. Not optimal but Free ;)) $BTC #Bitcoin #BTCUSDT #Trading #Crypto	long signal long binance btcusdtperp 4379657 20211001t000000z longterm signal appearing swingswiss daily chart bar close optimal free btc
2021-10-01 00:00:05				.0	.0	.0	.0	Bitcoin Daily: CoinEx to Shut Down China Business | https://t.co/j4aQweelpS https://t.co/5dmLPlFWfY	bitcoin daily coinex shut china business
2021-10-01 00:00:05				.0	.0	.0	.0	Investors are holding on to their bitcoins now more than ever. https://t.co/d3tAKS3NSh	investors holding bitcoins ever
2021-10-01 00:00:05				.0	.0	.0	.0	@GStein269 bitcoin	bitcoin
2021-10-01 00:00:06				.108	.104	.789	.0258	Next block fee rate: 1 sat/vByte Half hour fee rate: 1 sat/vByte Hour fee rate: 1 sat/vByte You can broadcast now your Bitcoin transaction #BroadcastNow	next block fee rate 1 satvbyte half hour fee rate 1 satvbyte hour fee rate 1 satvbyte broadcast bitcoin transaction
2021-10-01 00:00:06	3			.0	.0	.0	.0	ICYMI: U.S. Fed has no plans for a China-style crackdown on #cryptocurrencies, according to Jerome Powell https://t.co/GYxwGFwyhc #Bitcoin $BTC #cryptoregulation	icymi us fed plans chinastyle crackdown according jerome powell btc

BIBLIOGRAPHY

Aharon, D.Y., & Qadan, M. (2019) "Bitcoin and the Day-of-the-week Effect." *Finance Research Letters, 31,* 415–424.

AlNemer, H. A., Hkiri, B., & Khan, M. A. (2021) "Time-Varying Nexus between Investor Sentiment and Cryptocurrency Market: New Insights from a Wavelet Coherence Framework." *Journal of Risk and Financial Management, 14*(6), 275.

Anamika, C., M., & Subramaniam, S. (2021) "Does Sentiment Impact Cryptocurrency?" *Journal of Behavioral Finance,* 1-17.

Antweiler, W. & Frank, M. (2004) "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." *The Journal of Finance*, *59*(3), 1259–1294.

Baig, Ahmed, Blau, B. M., & Sabah. N. (2019) "Price Clustering and Sentiment in Bitcoin." *Finance Research Letters, 29:* 111–116.

Baur, D. G., Cahill, D., Godfrey, K., & Liu, Z. (2019) "Bitcoin Time-of-day, Day-of-week and Month-of-year Effects in Returns and Trading Volume." *Finance Research Letters, 31,* 78–92.

Caporale, G. M., & Plastun, A. (2019) "The Day of the Week Effect in the Cryptocurrency Market." *Finance Research Letters, 31*, 258-269.

Da, A., Engelberg, J, & Gao, P. (2015) "The Sum of All FEARS Investor Sentiment and Asset Prices." *Review of Financial Studies, 28*(1):1-32.

Décourt, R. F., Chohan, U. W., & Perugini, M. L. (2017). "Bitcoin Returns and the Monday Effect." *Horizontes Empresariales, 16*(2), 4-14

Durai, S. R. S., & Paul, S. (2018) "Calendar Anomaly and the Degree of Market Inefficiency of Bitcoin." *Madras School of Economics working paper*, (168).

Fama, E. F. (1970) “Efficient Capital Markets: A Review of Theory and Empirical Work.” *The Journal of Finance, 25*(2), 383-417.

Gao, X., Huang, W., & Wang, H. (2021) “Financial Twitter Sentiment on Bitcoin Return and High-Frequency Volatility.” *Virtual Economics, 4*(1), 7-18.

Glosten, R.L., Jagannathan, R., & Runkle, D.E. (1993). “On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks.” *Journal of Finance 48,* 1779–1801.

Gu, A. (2004) “The Reversing Weekend Effect: Evidence from the U.S. Equity Markets.” *Review of Quantitative Finance and Accounting, 22,* 5-14.

Guégan, D., & Renault, T. (2021) “Does Investor Sentiment on Social Media Provide Robust Information for Bitcoin Returns Predictability?” *Finance Research Letters, 38,* 101494.

Hutto, C & Gilbert, E. (2014) “Vader: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.” *Proceedings of the International AAAI Conference on Web and Social Media.*

Ibrahim, A. (2021) “Forecasting the Early Market Movement in Bitcoin Using Twitter's Sentiment Analysis: An Ensemble-Based Prediction Model.” *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 1-5. IEEE.

Jain, A.; Tripathi, S., Dwivedi, H.D., & Saxena, P. (2018) Forecasting Price of Cryptocurrencies Using Tweets Sentiment Analysis. *Proceedings of the Eleventh International Conference on Contemporary Computing (IC3).* Institute of Electrical and Electronics Engineers (IEEE), Noida, India, 2–4 August 2018; 1–7.

Jegadeesh, N & Wu, D. (2013). “Word Power: A New Approach for Content Analysis.” *Journal of Financial Economics,*110(3), 712–729.

Kaiser, L. (2019) “Seasonality in Cryptocurrencies.” *Finance Research Letters, 31,* 232–238.

Kapar, B., & Olmo, J. (2021) “Analysis of Bitcoin Prices Using Market and Sentiment Variables.” *The World Economy, 44*(1), 45-63

Kim S.-H. & Kim D. (2014) “Investor Sentiment from Internet Message Postings and The Predictability of Stock Returns.” *Journal of Economic Behavior and Organization, 107,* 708–729.

Kinateder, H., Papavassiliou, V.G., (2021) “Calendar Effects in Bitcoin Returns and Volatility.” *Finance Research Letters, 38,* 101420.

Kraaijeveld, O.; De Smedt, J. (2020) “The Predictive Power of Public Twitter Sentiment for Forecasting Cryptocurrency Prices.” *Journal of International Financial Markets, Institutions, and Money,* 65, 101188.

Kristoufek, L. (2013) “Bitcoin Meets Google Trends and Wikipedia: Quantifying the Relationship between Phenomena of the Internet Era.” *Scientific Reports, 3,* 3415.

Kurihara, Y., & Fukushima A. (2017) “The Market Efficiency of Bitcoin: A Weekly Anomaly Perspective.” *Journal of Applied Finance & Banking, 7*(3)*,* 57-64.

Li, T. R.; Chamrajnagar, A. S.; Fong, X. R.; Rizik, N. R.; Fu, F. (2019) “Sentiment-based Prediction of Alternative Cryptocurrency Price Fluctuations Using Gradient Boosting Tree Model.” *Frontier Physiology,* 7.

Lyócsa, S., Baumöhl, E., Výrost, T. Molnár. P. (2020) “Fear of the Coronavirus and the Stock Markets.”*Finance Research Letters, 36,* 101735.

Lyu, H.; Chen, L.; Wang, Y.; Luo, J. (2020) “Sense and Sensibility: Characterizing Social Media Users Regarding the Use of Controversial Terms for COVID-19.” *IEEE Transaction and Big Data,* 1.

Ma, D., Tanizaki, H. (2019) “The Day-of-the-week Effect on Bitcoin Return and Volatility.” *Research in International Business and Finance,* 49, 127–136.

Mohapatra, S.; Ahmed, N.; Alencar, P. (2019) “A Real-Time Cryptocurrency Price Prediction Platform Using Twitter Sentiments.” *arXiv:*2003.04967.

Nadarajah, S., Chu, J. (2017) “On the Inefficiency of Bitcoin.” *Economics Ketter,* 150, 6–9.

Naeem, M. A., Mbarki, I., and Shahzad, S. J. H. (2021) “Predictive Role of Online Investor Sentiment for Cryptocurrency Market: Evidence from Happiness and Fears.” *International Review of Economics and Finance, 73,* 496-514.

Qadan, M., Aharon, D. Y., and Eichel, R. (2021) “Seasonal and Calendar Effects and the Price Efficiency of Cryptocurrencies.”*Finance Research Letters*, 102354.

Öztürk, S. S., and Bilgiç, M. E. (2021) “Twitter and Bitcoin: Are the Most Influential Accounts Really Influential?.” *Applied Economics Letters*, 1-4.

Pano, T., Kashef, R. A. (2020) “Complete VADER-based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19.” *Big Data and Cognitive Computing, 4,* 33.

Ranasinghe, H., & Halgamuge, M. N. (2021) “Twitter Sentiment Data Analysis of User Behavior on Cryptocurrencies: Bitcoin and Ethereum.” *Analyzing Global Social Media Consumption,* 277-291, IGI Global.

Renault T. (2017) “Intraday Online Investor Sentiment and Return Patterns in the U.S. Stock Market.” *Journal of Banking and Finance, 84,* 25–40.

Rognone, L., Hyde, S. & Zhang, S. (2020) “News Sentiment in the Cryptocurrency Market: An Empirical Comparison with Forex.” *International Review of Financial Analysis, 69:* 101462.

Sadka, R & Scherbina A. (2007) “Analyst Disagreement, Mispricing, and Liquidity.” *The*

Journal of Finance, 6(5), 2367–2403.

Sailunaz, K., & Alhajj, R. (2019) "Emotion and Sentiment Analysis from Twitter Text." *Journal of Computational Science, 36*, 101003.

Stavroyiannis, S., Babalos, V., Bekiros, S., Lahmiri, S., & Salah G. (2019) "The High Frequency Multifractal Properties of Bitcoin." *Physica A: Statistical Mechanics and its Applications,* 520: 62–71.

Steyn, D. Greyling, T., Rossouw, S., & Mwamba, J. H. (2020) "Sentiment, Emotions and Stock Market Predictability in Developed and Emerging Markets." *No 502,GLO Discussion Paper Series from Global Labor Organization.*

Sun, Y., Kong, X., Chen, T., Su, H., Zeng, X., & Shen, Y. (2021) "Measuring Investor Sentiment of Cryptocurrency Market: Using Textual Analytics on Chain Node." *Procedia Computer Science,187,* 542-548.

Susana, D., Sreejith, S., & Kavisanmathi, J. K. (2020) "A Study on Calendar Anomalies in the Cryptocurrency Market." *International Working Conference on Transfer and Diffusion of IT,*166-177. Springer, Cham.

Tetlock, P. C. (2007) "Content to Investor Sentiment: The Role of Media in the Stock Market." *The Journal of Finance, 62*(3), 1139–1168.

Vega, C., (2006) "Stock Price Reaction to Public and Private Information." *Journal of Financial Economics, 82* (1), 103–133.

Yaya, O. S., & Ogbonna, E. A. (2019) "Do We Experience Day-of-the-week Effects in Returns and Volatility of Cryptocurrency?" *MPRA Paper No. 91429.*

Xie, P. (2021) "The Interplay between Investor Activity on Virtual Investment Community and the Trading Dynamics: Evidence from the Bitcoin Market."*Information Systems Frontiers,*

1-17

ACADEMIC VITA

CHEN-HAN LIU

cjl5932@psu.edu
LinkedIn: Chen-Han Liu

EDUCATION

The Pennsylvania State University, University Park, PA

- Smeal College of Business, Finance, B.S. — 2018 - 2022
- Schreyer Honors College — 2020 - 2022

AWARDS

The Evan Pugh Scholar Award — Apr. 2022

- Presented to juniors and seniors whose GPA is within the top 0.5%.

The Evan Pugh Scholar Award — Apr. 2021

- Presented to juniors and seniors whose GPA is within the top 0.5%.

The President's Sparks Award — Apr. 2020

- Presented to undergraduate students who have earned a 4.0 cumulative GPA based on at least 36 credits.

The President's Freshman Award — Apr. 2019

- Presented to undergraduate students who have earned a 4.0 cumulative GPA based on at least 12 credits.

7-time Dean's List Award

WORK EXPERIENCE

Investment Consultant and Accounting Intern — May 2021 - Aug. 2021

ABeat Semiconductor Ltd.

- Designed an optimal preferred stocks investment portfolio based on investments' historical performance, dividend yield, ratings, and board members' preferences.
- Contributed to the preparation of 2021 Q2 financial statements using Python and spreadsheet analysis.
- Devised a customized investment model approved by the CFO, using spreadsheet analysis and advanced VBA skills. The model is estimated to be valid until major changes are made to the investment direction or preferences.

Digital Finance Intern — Jun. 2020 - Jul. 2020

Bank SinoPac Ltd.

- Analyzed user feedback regarding the company's online banking system, DAWHO, using sentiment analysis.
- Assessed the potency of advertisement through traffic as a certified Google Ads and Google Analytics user.
- Devised the primary structure for the Search Engine Optimization (SEO) on Bank SinoPac's loans webpages

Marketing and Communications Intern — May 2019 - Aug. 2019

FOX Networks Group, Taiwan

- Performed daily tone analyses on feedback for shows on National Geographic and drafted reports on advice based on the results.
- Wrote press releases and managed social media pages for National Geographic's 50th Anniversary of the Moon Landing series.

Finance and Economics Instructor — Sep. 2019 - Aug. 2020

National Chi Nan University (Taiwan), Krirk University (Thailand) in partnership with Professor Vincent Ho

- Designed and led a year-long online workshop for 30+ undergraduate students with different backgrounds.
- Instructed finance, economics, symbolic logic, and statistics concepts using Python and Excel.

ACTIVITIES

World Vision International - Child Sponsor and Translator — Jan. 2015 - Present

Intercollegiate Taiwanese American Students Association - National Finance Assistant Director — Aug. 2018 - Jul. 2021

SKILLS

Technical Strengths — Python, Sentiment Analysis, Machine Learning, Data Analysis, Excel, VBA, Statistics

Languages — English, Mandarin, Taiwanese