Handelshøyskolen BI

GRA 19703 Master Thesis

Thesis Master of Science 100% - W

Predefinert informasjon

Startdato:	16-01-2022 09:00	**Termin:**	202210
Sluttdato:	01-07-2022 12:00	**Vurderingsform:**	Norsk 6-trinns skala (A-F)
Eksamensform:	T		
Flowkode:	202210||10936||IN00||W||T		
Intern sensor:	(Anonymisert)		

Deltaker

Navn:	Ellinor Giske Bakken og Veronika Høgset

Informasjon fra deltaker

Tittel *:	Bitcoin Returns - a Mood Killer?
Navn på veileder *:	Daniel Kim

Inneholder besvarelsen konfidensielt materiale?:	Nei	**Kan besvarelsen offentliggjøres?:**	Ja

Gruppe

Gruppenavn:	(Anonymisert)
Gruppenummer:	61
Andre medlemmer i gruppen:	

Bitcoin Returns – a Mood Killer?

Examination code and name:

GRA1970 Master Thesis

Authors:

Ellinor Giske Bakken and Veronika Høgset

Supervisor:

Daniel Kim

Programme:

Master of Science in Business with Major in Finance

Abstract

The thesis examines whether cryptocurrency returns may be used to explain sentiment expressed on social media. To do this, we have constructed a unique dataset consisting of posts on Twitter and Reddit mentioning Bitcoin and market data. The average sentiment scores on Twitter and Reddit are predicted with a linear regression model using Bitcoin returns.

Our analysis finds that Twitter users react positively to price increases and negatively to price falls. However, the explanatory power of the model is somewhat limited. In addition, we discover disparities in predicting power within the sample. When examining the proportion of positive and negative posts, we find that price movements impact the proportion of negative posts more than positive. On the other hand, we find no meaningful relationship between Bitcoin prices and sentiment expressed on Reddit. This thesis explores whether differences in risk preferences between Twitter and Reddit users might explain these findings.

Acknowledgments

First, we would thank our supervisor, assistant professor Daniel Kim of the Department of Finance at BI Norwegian Business School. Throughout our project, he has provided us with crucial advice and guidance. We would also thank our fellow students, Kristoffer Søderblom Jensen and Markus Lauvsland, at BI Norwegian Business School for their unconditional support throughout our two years of study at BI. Finally, we thank Tom Barr from WOCU for providing the World Unit Currency data sample for our regression analysis.

Table of Content

Table of Figures

1 Introduction

In 2008, “Bitcoin: A Peer-to-Peer Electronic Cash System” was published under the pseudonym Satoshi Nakamoto (Nakamoto, 2008). This paper introduced bitcoin as “a peer-to-peer version of electronic cash.” The following year, the first Bitcoin was created. In 2010 the first retail transaction was made, and it became tradable on exchanges (Hicks & Likos, 2022). From a low of $0.09 in 2010 to today’s price of approximately $42,000, Bitcoin has experienced a severe upswing (Edwards et al., 2022). More than 8,000 cryptocurrencies have been introduced in this period, such as Ethereum, Litecoin, and Monero (Hicks, 2020). Miners and coders have further developed the code behind bitcoin’s blockchain, making it adaptable for different uses.

The supply of Bitcoin is limited; only 21 million bitcoins can be mined. However, due to rounding operators in the Bitcoin database, the number of bitcoins will never reach this exact quantity. As of May 2022, 19 million bitcoins exist, and every 10 minutes, a new bitcoin is added to the supply (Hayes et al., 2022). The last Bitcoin is not expected to be generated before 2140. Due to a limited supply, Bitcoin is subject to deflation. While deflation is undesirable in traditional finance, it is positive for cryptocurrencies. Over the past decade, the consumer price index has risen by 28 percent, denominating that index in Bitcoin reveals a 99.9996 percent deflation, according to Bloomberg (Hajric, 2021).

Bitcoin was initially created as a means of exchange, but cryptocurrencies are often viewed as an investment vehicle. Unlike traditional assets, cryptocurrency has no apparent intrinsic value, and investors earn profit solely from price increases. It is difficult to identify the exact price drivers of cryptocurrencies, but the price depends on people’s perceptions and beliefs. Previous research has shown a significant relationship between cryptocurrency price and popularity (Saleh, 2018). For example, research suggests that Google Search Trends and Wikipedia affect the price of bitcoin (Kristoufek, 2013). Popularity can also be measured in social media activity. Overall, social media is “an important predictor of the future value of bitcoin” (Mai et al., 2018).

"Social media captures the wisdom of the crowd," according to Luo et al. (2013, p. 146). Generally, there are two types of investors: retail and institutions. Retail investors have primarily dominated the cryptocurrency markets (Subramaniam & Chakraborty, 2019). However, in the first quarter of 2021, institutional investors bought approximately the same amount of Bitcoin as retail investors (Dantes, 2021). Unlike institutional investors, retail investors express their opinions and create engagement among others on social media. In traditional finance, this is often called "talking your book". Because of a higher proportion of institutional investors, we believe that the current social media effect on prices is lower than what is found in previous research.

Past research has not addressed how the returns affect people's opinions and moods on social media. In this thesis, we want to examine how price movements affect the sentiment expressed on Twitter and Reddit. Do people's opinions about cryptocurrency become more favorable or adverse after a price movement? We pose the following central research question: *How do people react to cryptocurrency price movements on social media?* Before we test our research question, it is essential to understand and identify how the market behaves, what drives the price, and retail investors' risk behavior and tolerance.

To investigate our question of matter, we have collected a large and unique data sample containing Twitter and Reddit posts, Bitcoin prices, and other market data such as S&P 500, 3 Month Treasury Bill Rate, and World Currency Unit. Then, tweets and posts are subject to sentiment analysis. The outcome variables are used in a linear regression model to predict the sentiment on Twitter and Reddit in response to Bitcoin price swings with lags of up to 8 days. As a result, we can keep or reject the null hypothesis: *Cryptocurrency price movements do not affect social media activity.* Furthermore, we ran robustness checks to ensure our coefficients were stable predictors.

Our interest in blockchain technology and cryptocurrencies inspired us to write this thesis. We are curious about how the cryptocurrency market could grow so large, despite many in the financial industry being against it. Cryptocurrencies have not been covered in our mandatory classes; thus, we would like to take the opportunity to learn something new. Blockchain and cryptocurrencies represent a prodigious fintech innovation and conceivably a market shift. Inspired by the GameStop case and meme coins like Dogecoin, we focused on the relationship between social media and prices. We want to examine how returns affect social media indicators rather than analyze how social media indicators affect pricing as prior research has done.

The paper is organized as follows: In Section 2, we review previous literature, defining Bitcoin as an asset or currency and the relationship between social media and the cryptocurrency market. Section 3 presents the data used in our analysis and how it is collected. Further, Section 4 describes empirical methods, and Section 5 assesses the model's performance. Lastly, Section 6 discusses the implication and limitations of our study.

2 Theory

2.1 The Efficient Market Hypothesis

The Efficient Market Hypothesis proposed by Eugene Fama suggests that prices fully reflect all available information (Fama, 1969). The hypothesis is offered in three degrees: weak, semi-strong, and strong. The weak version implies that prices follow a random walk; hence, it is impossible to predict price movements. The concept of efficient markets was first applied to the stock market but has since extended to other assets like currencies and commodities (Evite, 2018). Therefore, the efficient market theory is relevant regardless of whether we consider cryptocurrency an investment asset or a currency.

Urquard (2016, p. 82) found that "[..] the inefficiency of Bitcoin is quite strong". However, he argues that the Bitcoin market will become more efficient as more investors trade. A follow-up study "shows for the first time that a power transformation of Bitcoin returns can be weakly efficient." (Nadarajah & Chu, 2017, p. 6). Tiwari et al. (2018) and Bariviera (2017) support the increased efficiency findings. On the other hand, Jiang et al. (2018) found no evidence of the bitcoin market becoming efficient. Unsurprisingly, Al-Yahyee et al. (2018) uncovered that gold, stocks, and currencies are less inefficient than Bitcoin.

If the strong or semi-strong version of the hypothesis applies to the cryptocurrency market, we would expect to see no differences in sentiment caused by price movements. The price should reflect all information; only new information could change people's opinion of cryptocurrencies. The weak form, or random walk hypothesis, implies that people could change their thoughts about cryptocurrencies following a significant price increase or decrease. However, the random walk hypothesis states that past prices cannot predict future prices, meaning people should not become optimistic or pessimistic based on only one price movement because it does not indicate future returns.

2.2 What Drives the Price?

A classification of cryptocurrency as an investment asset or currency is an important underlying feature of this thesis because it can say if or how people should react to price movements. The lack of regulatory classification challenges this operation; however, the European Central Bank is evident. Cryptocurrencies lack essential elements to be considered a currency (European Central Bank, 2021). Central banks do not back them, are accepted as a payment method only in a few places, and have no legal protection. Yermack (2015) argues that Bitcoin only fulfills the last attribute that money should have; it is a unit of account, a value store, and a medium of exchange. Disregardless of these features, or lack of features, it is not clear whether cryptocurrency behaves like an investment asset, a currency, or something else. The question is, what drives the price?

One may argue that cryptocurrencies, particularly Bitcoin, blur the boundaries between financial and monetary assets. However, people's actions often acknowledge the contribution of assets to the economy. According to Wolla (2018), the excitement over Bitcoin centered on purchasing it as a financial investment rather than an exchange medium. Glaser et al. (2014) support this and find that users interested in digital currencies are looking for an alternative investment vehicle rather than an alternative transaction mechanism. On the other hand, the Norwegian corporation Aker points out that financial access without a third party is the main reason for buying Bitcoin (Røkke, 2021). Furthermore, Bitcoin's daily exchange rate does not correlate with the US dollar or other dollar's exchange rates against major currencies or gold (Yermack, 2015). These findings suggest that Bitcoin resembles an investment vehicle.

Unlike equities, cryptocurrencies initially have no intrinsic value. However, Foley et al. (2019) argue that the illegal use of Bitcoin contributes to the fundamental value. The researchers claim that as much as 46 percent of bitcoin transactions from 2009 to 2016 were used for illegal activity. These findings sharply contrast those of more recent studies by Chainalysis (2021). They find that the illicit share of all cryptocurrency activity fell from 2.1 percent in 2019 to 0.34 percent in 2020. One reason for the fall is the increase in market capitalization. "[..] there are relatively fewer illegal users when bitcoin market capitalization is higher" (Foley et al., 2019, p. 1835). If illegal use no longer contributes to the fundamental value, something else must drive the price.

Some researchers show that cryptocurrencies have a low correlation and dependence on traditional financial assets like stocks (Bouri et al., 2020; Corbet et al., 2018; Tiwari et al., 2019). Low correlation may imply that the investors or price drivers are different. Liu and Tsyvinski (2018) argue that the returns can only be predicted using indications unique to the cryptocurrency market. On the other hand, new research has established the likelihood of financial market and cryptocurrency spillovers. Iyer (2022) found a significant increase in correlation between cryptocurrencies and S&P 500 from the pre-pandemic period to the post-pandemic period. "These findings suggest that crypto-assets may no longer be considered a fringe asset class and could potentially pose financial stability risks due to their extreme price volatility" (Iyer, 2022, p. 3).

Higher interest rates make government bonds more attractive compared to high-volatility and low-yield investments, and vice-versa. Cryptocurrencies' values plunged in early 2022, proving their position as a risky asset that investors dump in times of market stress. The decline was driven by the Federal Reserve's consideration of a faster interest rate increase (Ostroff, 2022). Corbet et al. (2017, p. 71) found that "[...] bitcoin volatility is influenced by both interest rate adjustments and QE announcement". New research supports these findings (Koutmos, 2019; Jareño et al., 2020). These findings are somewhat surprising given that cryptocurrencies are intended to be a decentralized currency independent of government intervention.

Bitcoin price can be closely related to its cost of production, according to Hayes (2017). The difficulty levels of Bitcoin's algorithms can speed up or slow down Bitcoin creation, changing its overall supply and its price. According to Ciaian et al. (2015), market forces such as Bitcoin supply and demand influence the price. As Bitcoin has grown in popularity, the impact of demand has increased.

The correlation between cryptocurrencies may drive the price. Bitcoin is the principal cryptocurrency, and it is therefore likely that Bitcoin will drive the price of other coins. Ajaz & Kumar (2019) state that the price of Bitcoin drives the price of Ethereum, Lite, and Dogecoin. Other researchers find that Bitcoin and Ethereum have a lagged relationship (Al-Jarrah et al., 2019; Katsiampa, 2019). However, both prices respond to important news, and the correlation may result from a common factor driving the price. Al-Jarrah et al. (2019) also find a relationship between Bitcoin, Monero, Dash, and Ripple.

In most countries, cryptocurrencies remain mostly unregulated and are frequently assumed to operate beyond the scope of national legislation. However, regulatory developments have significantly affected the price (Shanaev et al., 2020; Zeng et al., 2020). Auer & Claessenes (2018) show that broad cryptocurrency restrictions and their treatment in legislation have the most significant effect on the price. For example, when China prohibited all financial transactions involving cryptocurrency, the price of Bitcoin fell by 7 percent (Qin & Livni, 2021). On the other side, Wolk (2019) argues that the price of cryptocurrencies is determined by people's views and perceptions rather than by regulatory developments.

2.3 Research Question

As Liu and Tsyvinski (2018) stated, returns can only be forecasted with factors unique to cryptocurrency markets. The price levels depend on people's insight and opinions, which is an important feature of the crypto market, according to Wolk (2019). Therefore, analyzing the relationship between prices and social media indicators is essential. Mai et al. (2018) find that social media is an important indicator of future Bitcoin returns. On the other hand, Urquhart (2018) argues that investor attention has no significant predicting power of future prices, but realized volume and volatility affect investors' attention. Other researchers discovered the same causal relationship between stock returns and message board activity (Kim & Kim, 2014; Turmarkin & Whitelaw, 2001). Market information affects message board activity, not the other way around. These findings align with market efficiency; internet activity should not influence the price. Because Iyer (2022) shows an increase in correlation between S&P 500 and Bitcoin, these findings may have a transmissible effect on the crypto market.

Previous research which concludes that social media is an essential indicator of future Bitcoin returns could result from less common knowledge and smaller trading amounts.

Therefore, social media may now have a smaller effect. The findings of Ciaian et al. (2015) support this assumption, indicating that while Bitcoin was less well-known, online information had a more significant influence on Bitcoin price than in subsequent years due to Bitcoin’s establishment in the market.

Retail investors dominate the discussion on social media, and how retail investors behave on social media is an important aspect. If people discussing Bitcoin on social media have invested in it, their reaction to price movements will probably depend on their risk preference. Demographic factors affect people’s financial risk tolerance (FRT) and risk-taking behavior (FRB), according to Kannadhasan (2015). He analyses these terms by gender, age, marital status, income, occupation, and education. The study reveals that men are more risk-tolerant and risk-taking than women because women have more family responsibilities and lower lifetime earnings potential.

Furthermore, Kannadhasan (2015) finds the most significant differences in FRT and FRB between ages. The younger the investors are, the higher risk tolerance and increased risk behavior. An interesting finding in the study is that education help explains differences in FRT but not FRB. Retail investors with professional education avoid taking risks. Typically, they are individuals who use knowledge, among other means, to gain a competitive advantage when selecting investing opportunities and avoiding risky behavior. Further, occupation affects both FRT and FRB. A self-employed investor tends to have more risk-taking behavior and risk tolerance than salaried investors. On the other hand, income affects FRB but not FRT.

3 Data

3.1 Asset Selection

To test our hypothesis, we chose to focus on Bitcoin. We could have used other cryptocurrencies, but we believe Bitcoin represents the overall cryptocurrency market. Bitcoin has a market capitalization of 579 billion dollars per May 17, 2022, representing 43 percent of the overall market (Statistia, 2022). Figure 3.1.1 shows that market capitalization has decreased over time, but Bitcoin is still the dominant coin. In addition, we compared the age, gender, and nationality of Bitcoin owners to cryptocurrency holders in general and found no significant differences.

Figure 3.1.1: Bitcoin market capitalization as a percentage of total cryptocurrency market capitalization from March 2013 to May 2022.

Furthermore, we choose to focus on a broadly recognized cryptocurrency; data shows that 89 percent of U.S. adults have heard about Bitcoin (Wolfson, 2022). This recognition results in an enormous amount of Tweets and Reddit posts, which are beneficial for our statistical analysis. In addition, we choose Bitcoin because it is a stand-alone cryptocurrency, as opposed to ether and ripple, which are part of a larger network.

Bitcoin has a limited supply of 21 million coins and is therefore deflationary by nature (Hayes et al., 2022). Other coins with limited supply exist, such as Litecoin and Ripple, but these have much smaller market capitalization than Bitcoin. Coins with unlimited supply are, for example, Ethereum and Dogecoin. Every year, around 5.2 billion new Dogecoins are created, implying that the investor's value will diminish over time (Leech, 2021). Cryptocurrencies that are hard-capped have a greater chance of preserving their value. Coins with unlimited supply are short-term investments, not long-term investments, commonly used as a pump and dump scheme. We want to examine a coin used as both short-term and long-term investment, such as Bitcoin, and rule out coins with an unlimited supply.

Because we will look at the sentiment scores, we want to examine a cryptocurrency that people are somewhat neutral to in the first place. The hype on social media drives the price of meme coins like Dogecoin and Shiba Inu, but the coins lack support in parts of the cryptocurrency community. An article by Bloomberg states that many people in the crypto space are de facto short Dogecoin (Weisenthal, 2021). These coins are frequently seen as an online joke because they lack certain features compared to originals like Bitcoin. Meme coins have no intrinsic value and offer very little to their users, implying that they are not long-term investments.

3.2 Returns

In the following analysis, we use returns as regressor variables. Returns are calculated on a logarithmic basis and derived using Equation 3.2.1.

$$Return_t = \ln\left(\frac{P_t - P_{t-1}}{P_{t-1}}\right) = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{3.2.1}$$

Natural logarithmic returns are used instead of simple returns to secure the validity of our analysis. One of the assumptions underlying the classical linear regression model is that the residuals are homoscedastic, i.e., the residuals are constant and do not vary over time (Konermann, 2021). Stock markets, and cryptocurrency markets, are often examples of where conditional heteroskedasticity is present (Hayes et al., 2022). Using natural logarithmic values avoids the potential violation of this assumption.

3.3 Twitter

Twitter is a social media platform founded in 2006. Since then, Twitter has expanded to 217 million monetizable daily active users (Twitter, 2022), making it one of the most popular social networking sites. Users on Twitter may create tweets, which are short messages of 280 characters or less, possible to share with the whole world. Users use the character "#" or hashtag with the following topic word to identify tweets by themes, making it easier for users to follow what they are interested in reading. Examples on hashtag can be: #Finance, #MondayMotivation and #HowToGetAnAonMasterThesis.

3.3 Reddit

Reddit is a social media platform where users can share posts, comment, and vote on other posts. It consists of different communities, called Subreddits, where people discuss subjects related to a specific theme or category. Examples of Subreddits are r/Music, r/Norway, r/Wallstreetbets, and r/Bitcoin. In total, there are 130 000 active Subreddits or communities. Reddit has approximately 52 million daily and 430 million monthly users (Todorov, 2022). On its homepage, Reddit writes that its platform "creates and catalyzes culture – a single comment can spark a global movement" (Reddit, 2022).

In 2021, Reddit, specifically r/Wallstreetbets, received much attention. Dozens of retail traders found each other on the subreddit, which now has over 12 million users (Reddit, 2022). They orchestrated a giant short squeeze in GameStop by motivating each other to buy shares and call options. The results were a 1,500 percent increase in stock price and significant losses for short-selling hedge funds (Choudhury, 2021). The case of GameStop is an excellent example of how technology can enhance individual impacts in the markets. Some of the same faddish dynamics have benefited the rise in the cryptocurrency market (Regnier, 2021).

3.4 Data Collection

We downloaded Bitcoin prices from Yahoo Finance from January 1, 2020, to March 1, 2022, as we want as new a set of data as possible. In addition, we choose this timeframe to see if the coefficient estimates behave any differently from the months when COVID-19 was at its worst, before and after. Due to computation time, we had to restrict the timeframe.

To collect data from Twitter, we installed and utilized the snscrape Python library (snscrape, 2022). It scrapes Twitter for tweets with certain words and returns the tweet, as well as the username, number of followers, likes, and retweets. Because Bitcoin is extensively discussed on Twitter, collecting data was time-consuming. It took around six days to download about 6.54 million tweets mentioning Bitcoin.

Scraping Reddit is more challenging than scraping Twitter since Reddit's API prevents such extensive data collecting. However, several empirical studies have used *Pushshift* to circumvent these restrictions. "Pushshift's Reddit dataset is updated in real-time and includes historical data back to Reddit's inception" (Baumgartner et al., 2020). While all posts are available through Pushshift, the historical database does not update with changes, which means we miss updated data about scores and the number of comments. It ultimately limits our ability to weight the sentiment score by popularity and split the dataset to check for significant differences. In addition, deleted posts on Reddit are not removed from Pushshift.

The possibility of the database going down one day was another challenge with Pushshift. It was a well-known problem that others had mentioned on the internet, which we encountered when we first started scraping Reddit through the PushshiftAPI in Python. However, determining whether our code or the database caused the problem was challenging. Additionally, the issue may have resulted in some missing values in our dataset. In September 2021, we missed data for a total of seven days. In total, we collected approximately 236 thousand posts from the subreddit r/Bitcoin and the voting scores and number of comments.

3.5 Variable Construction

To conduct sentiment analysis, we use the Python library VADER, Valence Aware Dictionary, and sEntiment Reasoner (Hutto & Gilbert, 2014). VADER is a lexicon and rule-based sentiment analysis tool developed specifically for social media analysis (Hutto, 2021). Using our sample from Twitter and Reddit as input, VADER returns the positive sentiment score, negative sentiment score, neutral sentiment score, and the compounded score. The positive, negative and neutral sentiment scores are all ratios of the percentage of text that fall into each classification, adding up to 1. Text can have a proportion of positive, negative, and neutral scores. The reason is that VADER categorizes each lexical item in the text.

The compounded score is a weighted composite score normalized between -1 and 1, where -1 is most pessimistic and 1 is most favorable. Most researchers have used the compounded score in their analysis (Pano & Kashef, 2020; Kim. et al., 2016, Bjørgan et al., 2021). These scores can classify the tweet or post as either negative, neutral or positive. A compounded score equal to or greater than 0.05 is often characterized as positive sentiment, while a score less than -0.05 is considered negative (Britzolakis et al., 2020). Text with a compound score between those is considered neutral. Table 3.5.1 shows examples of tweets and posts from our dataset and their estimated sentiment score.

Table 3.5.1: Examples of tweets and their compound score, negative sentiment score, neutral sentiment score, and positive sentiment score.

Username	Tweet	Compound	Negative	Neutral	Positive
elonmusk	BTC (Bitcoin) is an anagram of TBC (The Boring Company) What a coincidence!	-0.38	0.19	0.81	0.00
AlexSaunders-AU	I’ve seen plenty of milestones since 2012. But something about this feels very special. Like global adoption is inevitable. Like $BTC value is heading for the trillions. Like 80%+ corrections are in the past. Like we are truly winning. Like #Bitcoin was the key to change. ✅ https://t.co/RgvhPoHseO	0.99	0.00	0.54	0.46
AlgoTrader-Mack	FUCK FUCK FUCK FUCK FUCK FUCK FUCK FUCK FUCK FUCK FUCK FUCK FUCK FUCK FUCK FUCK #btc #bitcoin $btc $btcusd #crypto #cryptocurrency https://t.co/FDI33EOBjH	-0.99	0.91	0.09	0.00

Furthermore, we calculated the mean compounded score for each day because our independent and additional variables were daily. We also calculated the proportion of total daily tweets or posts containing some negative, neutral, and positive lexical items, respectively. We primarily used MATLAB in the process of formatting and cleaning the dataset, in addition to computing variables mentioned above.

3.6 Additional Variables

Other variables, in addition to the original regression variables, were implemented in our analysis to check whether they would change the predicting power of the independent variables. We chose these variables partly based on our discussion in

Section 2.2 and partly because we thought they could have a significant relationship to the dependent variable. These included S&P 500, SSE Composite Index, 3 Month Treasury Bill Rate, 10 Year Treasury Bill Rate, WOCU, and SPCommodities Index. Of these, we chose to focus on the three most engrossing variables in our robustness checks, i.e., variables that had some significant effect on the sentiment score or have not been tested before. These were the S&P 500, 3 Month Treasury Bill Rate and WOCU.

As Iyver (2022) points out, there is a significant rise in the correlation between the returns of cryptocurrencies and the S&P 500, implying an increased correlation between stock market returns and Bitcoin sentiment scores. We chose three months, not ten years, Treasury Bill Rates because we saw a significant effect of the former and not the latter. As previously mentioned, Corbet et al. (2017) found that Bitcoin returns were affected by changes in the interest rate, meaning that interest rates could affect sentiment scores. We collected S&P500 and Treasury Bill Rates from *Global Financial Data* for the same period as cryptocurrency returns.

The last variable is the WOCU World Currency Unit, a reference quotation or “global currency” produced by WOCU Limited, a UK company (WOCU, n.d.). It is derived from a weighted basket of fiat currency pairs, covering the World’s top 20 economies by Gross Domestic Product (GDP). The WOCU reflects the market reality of a basket of real-time currency rates predicated on the evolving economic importance of its constituent currencies, representing around 80 percent of global GDP. We were provided the WOCU World Currency Unit after contacting the company.

4 Methodology

4.1 Sentiment Analysis

Sentiment analysis, often known as opinion mining, is a natural language processing (NLP) technique for detecting and categorizing the emotional tone of a text (Tonkin, 2016). NLP is a branch of artificial intelligence that combines computer science and computational linguistics (SAS, n.d.). We apply sentiment analysis to a vast collection of tweets and posts on Reddit to get a sense of how people feel about cryptocurrencies.

As previously mentioned in Section 3, we used the Python module VADER to perform the sentiment analysis. Hutto & Gilbert (2014) state that "[..] the VADER lexicon performs exceptionally well in the social media domain" and "[..] VADER performs as well as individual human raters". We find support for these findings in Saad & Yang (2019) and Bonta et al. (2019). Several papers employ this analyzing tool (Borg & Boldt, 2020; Park & Seo, 2018; Pano & Kashef, 2020; Abraham et al., 2018). VADER indicates if sentiment is positive or negative and to which degree. In addition, it considers degree modifiers, punctuation, word shape, slang, acronyms, and emojis to alter the sentiment intensity. Furthermore, VADER does not suffer from the conventional speed-performance tradeoff.

4.2 Linear Regression

We use a simple linear regression model for our analysis. The daily logarithmic Bitcoin returns are the independent variable with up to 8 lags. The sentiment scores for both Twitter and Reddit are dependent variables, including the mean compound score and the proportion of negative, positive score, and neutral posts. All linear regression models follow the same logic as described in Equation 4.2.1. The coefficients are estimated in Stata using ordinary least squares regression (OLS).

$$CompoundScore_t^{TwitterBTC} = \beta_0 + \beta_{1,BTC} * ReturnBTC_t + \beta_{2,BTC} * ReturnBTC_{t-1} + \beta_{3,BTC} * ReturnBTC_{t-2} + \beta_{4,BTC} * ReturnBTC_{t-3} + \beta_{5,BTC} * ReturnBTC_{t-4} + \beta_{6,BTC} * ReturnBTC_{t-5} + \beta_{7,BTC} * ReturnBTC_{t-6} + \beta_{8,BTC} * ReturnBTC_{t-7} + \beta_{9,BTC} * ReturnBTC_{t-8} + \varepsilon_t \tag{4.2.1}$$

An essential aspect of why we use independent variables at time t, the same as our dependent variable, is that social media responds very quickly to price changes. Other studies have often used the sentiment score the day before to predict prices. However, it is reasonable to believe that people respond somewhat immediately since the market is open 24 hours a day.

The linear regression model builds on five underlying assumptions. To obtain an appropriate model, we must first determine if the five assumptions hold (Brooks, 2019). We will follow the same technique for all regressions but only comment on the results from Equation 4.2.1. The first assumption is that the residuals have zero mean. The mean of the residuals in our model is 1.87e-11, approximately equal to zero.

The second assumption is that the variance of the residuals is constant and has finite overall values of X_t. If they are constant, they are homoscedastic. By producing a scatterplot of the residuals on the Y-axis and the predicted values of the dependent variable on the X-axis (Residuals-Versus-Fitted Plot), we can see if the residuals are concentrated around zero, which they are. The Residuals-Versus-Fitted Plot is found in the Appendix. We can also check for heteroskedasticity by using the Breus-Pegen and Cook-Weisberg tests. The p-value is greater than 0.10 for both tests, meaning we cannot reject the null hypothesis about homoscedasticity.

Third, residuals should be independent of one another, i.e., not autocorrelated. We reject the null hypothesis of no serial autocorrelation in the Durbin Watson test and find that the residuals have a positive correlation. OLS may no longer be a minimal variance estimator. To correct the autocorrelation problem, we use the Newey command in Stata. The Newey-West estimator adjusts the standard errors to correct for heteroscedasticity in the residuals (Brooks, 2019). Because Newey generates the same coefficient as in OLS, our predictions and R-square would be unaffected. If we want to know how much uncertainty there is around the predictions, we should examine the results when using Newey-West standard errors. As a result, the standard errors, t-statistics, and p-values changed marginally compared to the simple OLS model.

The fourth assumption states that no relationship exists between the residuals and corresponding X variables. The correlation table in the Appendix shows that the error terms have zero relationships to the lagged variables.

The last assumption says that the error term is normally distributed. One way to test for normality is the Shapiro-Wilk test, where the null hypothesis is that the variable has a normal distribution. The results show a p-value of zero, meaning the assumption does not hold for our residuals. Another way to test for normality is to use the skewness and kurtosis test, where a normal distribution is not skewed. Here we also obtain a p-value equal to zero. However, we believe that some extreme residuals cause the rejection of the null hypothesis. A plot in the Appendix shows some outliers around the start of the Covid-19 pandemic and 2021.

It is not obvious what one should do to solve this validation error but using dummy variables or another strategy to eliminate such observations is one way to increase the probability of error normality. On the other hand, many econometricians claim that using dummy variables to eliminate outlying residuals may artificially improve the model's features, thus manipulating the findings (Brooks, 2019). Even if assumption five does not hold, the OLS estimators are the "Best Linear Unbiased Estimators" (BLUE).

To compare and evaluate the predicting power of our model, we primarily use the R^2 and adjusted R^2. The former indicates how effectively our model can predict the values of the dependent variable in percentage terms.

5 Results

Below, we will present and discuss the results of our estimation of sentiment scores. First, we introduce the results from the simple linear regression model presented in Section 4.2. Second, we discuss the implications of our results, substantiated by financial theory and previous research. Lastly, we analyze how results could change when adding additional variables or limiting the time horizon.

5.1 Exploring the Impact of Returns on Sentiment Scores

First, we look at the impact of returns on average Twitter sentiment scores. We specify a classical linear regression model according to Equation 4.2 and use Newey West standard errors since the data is subject to autocorrelation. Table 5.1.1 contains the results of the regression.

TABLE 5.1.1
LINEAR REGRESSION COEFFICIENTS, EQUATION 4.2.1

BTC Return lag	Twitter Mean Compound Score $_{BTC, t}$
BTC Return	0.00116*** (0.00023)
Lag 1	0.00128*** (0.00029)
Lag 2	0.00068*** (0.00033)
Lag 3	0.00085*** (0.00027)
Lag 4	0.00106*** (0.00026)
Lag 5	0.00077*** (0.00025)
Lag 6	0.00093*** (0.00027)
Lag 7	0.00079*** (0.00027)
Lag 8	0.00036 (0.00027)
Constant	0.0990 (0.00123)
Observations	783
R^2	0.1443
Adjusted R^2	0.1344
F-statistics	8.48*** (df: 9; 773)

*Notes: Parameters of linear regression, with standard deviation in parenthesis. *10%, **5% and ***1% significance level.*

All independent variables, except the last lag, are significant at a 1% level. There is a positive relationship between Bitcoin returns and sentiment expressed on Twitter, which means that people, on average, react positively to price increases and negatively to a fall in prices. In other words, we reject our null hypothesis and conclude that returns, up to seven lags, affect sentiment scores. The average sentiment score is not directly affected by the number of tweets in a day. Hence, our results only indicate if the people who tweet are positive or negative and not the number of people who have a positive or negative view.

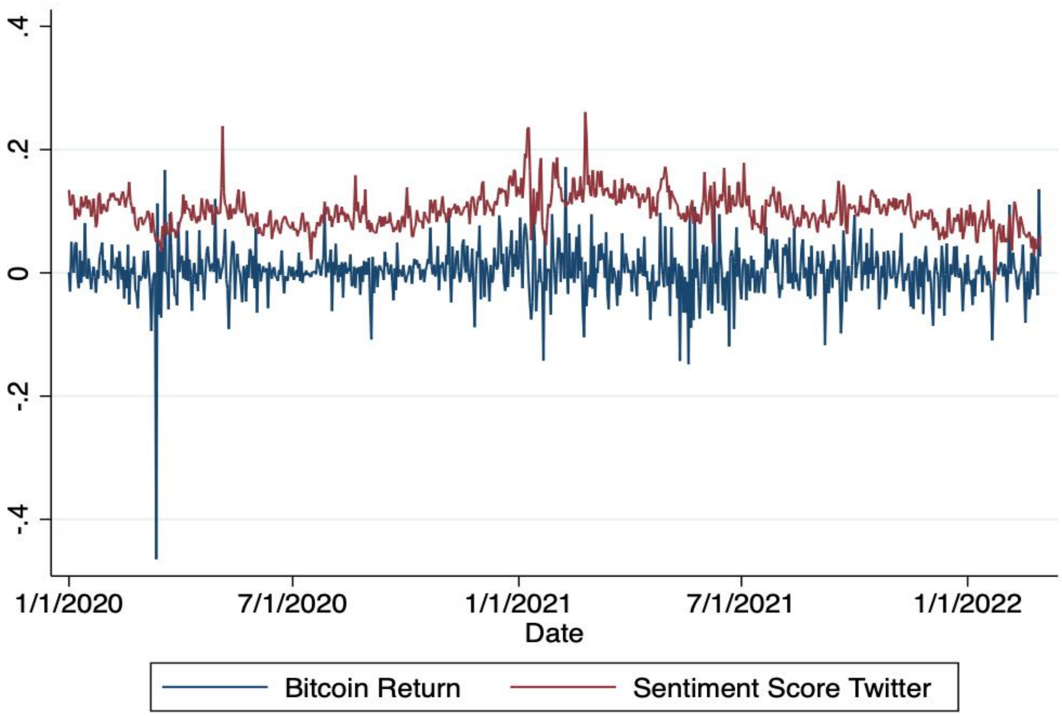
Notably, the adjusted R^2 for the model is only 14.43 percent, meaning that the model explains only a small part of sentiment scores. R^2 assumes that all independent variables affect the model's performance. As a result, the number of independent variables tends to increase R^2, while Adjusted R^2 only looks at the variables that affect the predicting power of the model. When adding another lag of the independent variable, we saw a decrease in Adjusted R^2.

As discussed in Section 2, these findings imply that the cryptocurrency market is inefficient and that Twitter users view Bitcoin as an investment rather than a medium of exchange. We would anticipate price movements to have smaller effect on sentiment if people thought of Bitcoin as a currency. Assuming that those who tweet are also invested in Bitcoin, a positive relationship between returns and sentiment indicates that investors are risk-averse. Risk is often defined as price volatility, and risk-averse investors will only take on additional risk if it is warranted by the potential for higher returns. Hence, they could react positively to price increases because it indicates a potential for higher returns.

Regulation news impacts Bitcoin returns, as discussed in Section 2.2. People may react to changes in regulations rather than price movements, resulting in a significant relationship between returns and sentiment scores. It is also possible that changes in regulations will have a more significant influence on sentiment scores than changes in returns. Furthermore, the correlation between S&P 500 and Bitcoin returns is 36 percent in our dataset, which could imply that stock returns indirectly affect sentiment score. We will further discuss this in Section 5.3. Since returns only explain some of the sentiment scores, there must be something else influencing it. The overall world economic situation might affect the sentiment score. Interest rates, GDP, and the unemployment rate could be indicators of the economic situation. We will look at how interest rates affect sentiment scores in Section 5.3.

Figure 5.1.1 plot the Bitcoin returns and the mean sentiment score of all tweets mentioning Bitcoin. The graph indicates that in times of large price movements, the sentiment scores vary more than when the market is less volatile. Bitcoin returns and average sentiment scores have a correlation of 16 percent. One of the lowest average sentiment scores appeared when the Bitcoin price fell by almost 50 percent on March 12, 2020, following the coronavirus outbreak.

Figure 5.1.1 Bitcoin returns and mean expressed sentiment score of tweets containing "Bitcoin" from January 1, 2020, to March 1, 2022.



Further, we look at how returns affect the percentage of tweets containing some positive or negative item in Table 5.1.2. For the negative proportion, all lags are significant, whereas all up to lag four are significant at a 1% level. These findings imply that when the return of Bitcoin decreases, the proportion of negative tweets increases. A critical remark is that when the return of Bitcoin increases, the proportion of negativity decreases, which means that people who dislike Bitcoin go silent or a small proportion goes positive.

For the positive proportion, there are less significant values. Only the first, fourth and fifth lag is significant at 10%, and the sixth at 5%. Indicating that only a small proportion of positive tweets decreases when the return of Bitcoin decreases and vice versa. Meaning, that Bitcoin lovers stay constant even though the price falls.

TABLE 5.1.2
LINEAR REGRESSION COEFFICIENTS, PROPORTION OF NEGATIVE AND POSITIVE TWEETS

BTC Return lag	Negative tweets in % of total $_{BTC, t}$	Positive tweets in % of total $_{BTC, t}$
BTC Return	-0.000253***	0.000117*
	(0.00004)	(0.00007)
Lag 1	-0.000283***	0.000104
	(0.00004)	(0.00008)
Lag 2	-0.000156***	0.000015
	(0.00005)	(0.00008)
Lag 3	-0.000159***	0.000073
	(0.00005)	(0.00007)
Lag 4	-0.000167***	0.000129*
	(0.00004)	(0.00007)
Lag 5	-0.000085**	0.000134*
	(0.00004)	(0.00007)
Lag 6	-0.000086**	0.000196**
	(0.00003)	(0.00007)
Lag 7	-0.000116***	0.000106
	(0.00004)	(0.00007)
Lag 8	-0.000065*	0.000061
	(0.00004)	(0.00008)
Constant	0.034249	0.061368
	(0.00022)	(0.00037)
Observations	783	783
R^2	0.1643	0.0324
Adjusted R^2	0.1536	0.0212
F-statistics	11.08*** (df: 9; 773)	1.55 (df: 9; 773)

*Notes: Parameters of linear regression, with standard deviation in parenthesis. *10%, **5% and ***1% significance level.*

The adjusted R^2 for the proportion of negative tweets is 15.36 percent, slightly higher than what is found in Table 5.1.1. However, the adjusted R^2 for the proportion of positive tweets is relatively low at 2.12 percent, which means that the significant values can be due to randomness as the model only explain a low portion of the variation of the dependent value.

Finally, we will look at the impact of returns on sentiment scores expressed on Reddit. When running the corresponding equation from 4.2.1, we found that Bitcoin returns have no significant effect on the average sentiment on Reddit, as seen in Table 5.1.3. Both R^2 and Adjusted R^2 are approximately equal to zero, meaning that the returns cannot explain the average sentiment score in a day.

TABLE 5.1.3
LINEAR REGRESSION COEFFICIENTS, EQUATION
REDDIT

BTC Return lag	Reddit Mean Compound Score $_{BTC, t}$
BTC Return	-0.00033 (0.00029)
Lag 1	-0.00013 (0.00024)
Lag 2	-0.00012 (0.00023)
Lag 3	-0.00026 (0.00026)
Lag 4	0.00025 (0.00024)
Lag 5	0.00019 (0.00028)
Lag 6	-0.00014 (0.00027)
Lag 7	0.00012 (0.00028)
Lag 8	0.00002 (0.00027)
Constant	0.07918 (0.00129)
Observations	783
R^2	0.005
Adjusted R^2	0.000
F-statistics	0.47 (df: 9; 773)

*Notes: Parameters of linear regression, with standard deviation in parenthesis. *10%, **5% and ***1% significance level.*

5.2 Differences between Twitter and Reddit

The difference in results surprised us because we expect Reddit and Twitter users to behave similarly. As we will further discuss in Section 6.1, non-lexical items such as pictures and GIFs in the post are not captured in our sentiment analysis. As a result, the overall expressed sentiment likely differs from the sentiment in the text. Therefore, one should interpret the results with caution. However, in the next session, we will address some plausible explanations for why individuals react differently on Twitter and Reddit.

5.2.1 Risk Tolerance and Behavior

As mentioned in Section 2.2, a classification of cryptocurrencies as either an investment asset or currency might have an effect on how people react to price movements. If people perceive cryptocurrencies as a means of exchange and acquire crypto for their capabilities, price movements should have less impact on the general opinions and sentiment. On the other hand, if the purpose of buying cryptocurrency is to make money, price movements should have an impact on sentiment. Maybe users on Reddit do not see Bitcoin as an investment in the same way that users on Twitter do. Suggesting that people on Reddit are less risk-averse. Could this indicate that Reddit users value Bitcoin as an exchange medium?

While differences in perceptions can be challenging to observe, we can observe differences in demographics. Table 5.2.1 shows the difference in demographics between Twitter and Reddit users. Differences in demographics may cause variation in risk preferences and risk behavior. These differences are for the average Twitter and Reddit user, not necessarily the differences between people who tweet about Bitcoin and post on r/Bitcoin. It is unreasonable to believe that one of the demographic disparities makes the results differ, but a combination of different characteristics may have some effect.

TABLE 5.2.1
TWITTER AND REDDIT DEMOGRAPHICS

	Twitter	Reddit
Age		
13-17	7%	7%*
18-24	17%	28%*
25-34	39%	30%*
35-49	21%	25%*
50+	17%	10%*
Gender	44%	36%
Female	56%	64%
Male		
Annual household income		
Under $30,000	23%	30%
$30,000-75,000	36%	34%
Over $75,000	41%	35%
Political preference		
Liberate	37%**	43%
Moderate	30%**	38%
Conservative	33%**	19%
Nationality	United States (26%) Japan (15%) United Kingdom (6%)	United States (47%) United Kingdom (8%) Canada (7%)

** Only Reddit users in the United States.*
*** Based on the percentage of people who place their own ideology on an 11-point scale. 0-4 are assumed to be liberal, 5-6 to be moderate, and 7-10 to be conservative.*

First, we observe that more younger people use Reddit than Twitter (Statistia, 2022; Pew Research Center, 2016). Are younger people less prone to changes in prices? One might assume that young people have less money and are consequently more concerned about returns. However, as mentioned by Kannadashan (2015), younger investors have a significantly higher risk tolerance and behavior than older people. These findings can be driven by the fear of missing out (FOMO), leading young people to invest money they cannot afford in risky assets that have already experienced rapid growth and not think about their long-term strategy.

According to a survey by Interactive Investor, almost half of young Brits who chose cryptocurrency as their first investment use loans to put bets on these assets (Sawhney, 2021). It can then seem like young people do not care if the assets are risky or not. Young investors will often gain knowledge from social media or blindly follow other investors' recommendations instead of watching market changes (Amboala, 2018). When young people follow social norms rather than deviating from them, they feel safe (Amboala, 2018). Losing in a group can be easier to forget and accept. It could help explain why the sentiment score on Reddit is unaffected by Bitcoin returns, as seen in Table 5.1.3.

Age correlates with education and professional experience. Younger people are generally less educated and have shorter career history. However, the number of people taking a bachelor's degree or higher keeps increasing (Nietzel, 2021), and data shows that most people graduate at the age of 22 to 24 (Brunner, n.d.). It was challenging to find comparable numbers on the education of users of Twitter and Reddit, and the age is, therefore, the most appropriate approximate. Following the same logic as in the paragraph above, it is reasonable to believe that individuals with higher education care more about the price and know the implications of a price movement.

Cryptocurrency owners are, in general, more educated than the average population. According to The Bank of International Settlement, the higher an individual's education, the higher the probability of owning a cryptocurrency (BIS, 2021). People often view cryptocurrencies as a more complex investment opportunity, and it is not surprising that the owners are generally more educated. However, there are considerable differences within the group of cryptocurrency owners, and Bitcoin owners have less education than people owning, for example, Ripple, Ethereum, and Bitcoin Cash (BIS, 2021). People with higher education might express more negative opinions about a volatile investment such as Bitcoin because they are risk-averse. Assuming that people on Twitter have higher education because they are older supports our findings in Table 5.1.2 that the percentage of negative tweets is most affected by returns.

Second, fewer females use Reddit (Statistia, 2022; Statistia, 2022). Jianakoplos and Bernasek (2007) state that women have a higher risk aversion than single males when making financial decisions. This finding should imply that women are more concerned with returns. Risk equals price volatility when investing, implying that women should react negatively to significant price movements, particularly substantial falls. It supports our findings in Section 5.1 that the percentage of negative tweets increases when the price decreases, and vice versa.

Twitter users are slightly wealthier than Reddit users (Wojcik & Hughes, 2019; Pew Research Center, 2016), which might be because of the many young people on Reddit. As previously mentioned, young people are less risk-averse because they often use money they do not afford to lose. In addition, young people might lack the expertise or education to recognize the risk of price movements. Kannadasan (2015) shows that there is a positive relationship between income and risk behavior. If we do not assume that all individuals discussing Bitcoin on social media are invested in it, then a higher proportion of Twitter users than Reddit users should have invested in a risky asset such as Bitcoin. People owning Bitcoin are more likely to respond to price movements, which could support our findings in Tables 5.1.1 and 5.1.3.

There are some differences in political preferences between users of Twitter versus Reddit (Pew Research Center, 2018; Pew Research Center, 2016). A higher percentage of Twitter users are conservative compared to Reddit users. According to Jiang et al. (2016), political donations and forecasting behavior are linked. The study reveals that people donating to the republican party are more conservative in their forecasting behavior. As a result, conservative analysts are likely to be more rational and better informed than the average analyst. These results may suggest that people on Reddit are less rational and maybe overconfident because many users are liberal. Hence, they would not incorporate adverse price movements into their sentiment.

Lastly, Americans dominate Reddit; 50 percent of Reddit users are Americans compared to 25 percent on Twitter (Statistia, 2022; Statistia, 2022). Are people from Twitter more risk-averse because of the significant cultural differences? A study by Hens et al. (2011) discovered that even when variables like inflation rates and wealth are taken into consideration, the cultural background has an impact on investing risk behavior. According to their research, investors from the United States are more "ego-traders" than most countries in Europe. Meaning that they can never go wrong with their investing decisions, indicating that they will stick to their trading strategy regardless of whether the price of Bitcoin rises or falls. Twitter, which has more differences in cultural areas, reacts more to price changes and is, therefore, more risk-averse. These arguments support our findings in Table 5.1.1 and Table 5.1.3.

5.2.2 The Case of GameStop

We mentioned in Section 3.3 that Reddit received much attention when it orchestrated a giant short squeeze in GameStop. GameStop can be compared to cryptocurrencies because it is a meme stock with little fundamental value. “The volatility tied more closely to non-fundamental trading, social media influence and other factors” (Yahoo Finance, 2022). Would our analysis be any different if we look at stocks like GameStop instead of Bitcoin?

Overall, we would expect to see a significant relationship between GameStop returns and sentiment expressed on r/Wallstreetbets. Initially, the investing app Robinhood played a critical role when the Reddit posers saw an opportunity to make money while also giving a jab to Wall Street and hedge funds (Gonzales & Priest, 2021). Welch (2021) finds that Robinhood investors bought individual stocks after large price movements and that the entertainment value of their trading is sometimes more important than, say, risk and return. This might imply that as the entertainment increases, that is, the price and return, the more positive the sentiment becomes. Indicating a positive relationship between returns and sentiment.

We anticipate seeing a stronger relationship between sentiment and GameStop returns than Bitcoin return. The whole point of buying GameStop was to boost the price, both for personal returns and because higher prices would damage short-selling hedge funds even more.

As previously stated, Reddit primarily consists of young people. They often rely more on the information they gather from one another than on their own research. While GameStop’s price was low, the return had probably an impact on how people responded on Reddit. However, once the price began to rise dramatically, more people wanted to ride the wave and “talked their book” on Reddit.

5.3 Robustness Test

We will examine how our main regression coefficient estimates behave when the additional variables discussed in Section 4.6 are included. A critical remark is that these variables do not trade every day, such as cryptocurrencies. As a result, we exclude these "non-trading" days and end up with 535 observations in our regression. The outcome of a regression with three additional variables are shown in Table 5.3.1.

TABLE 5.3.1
LINEAR REGRESSION COEFFICIENTS ADDITIONAL VARIABLES

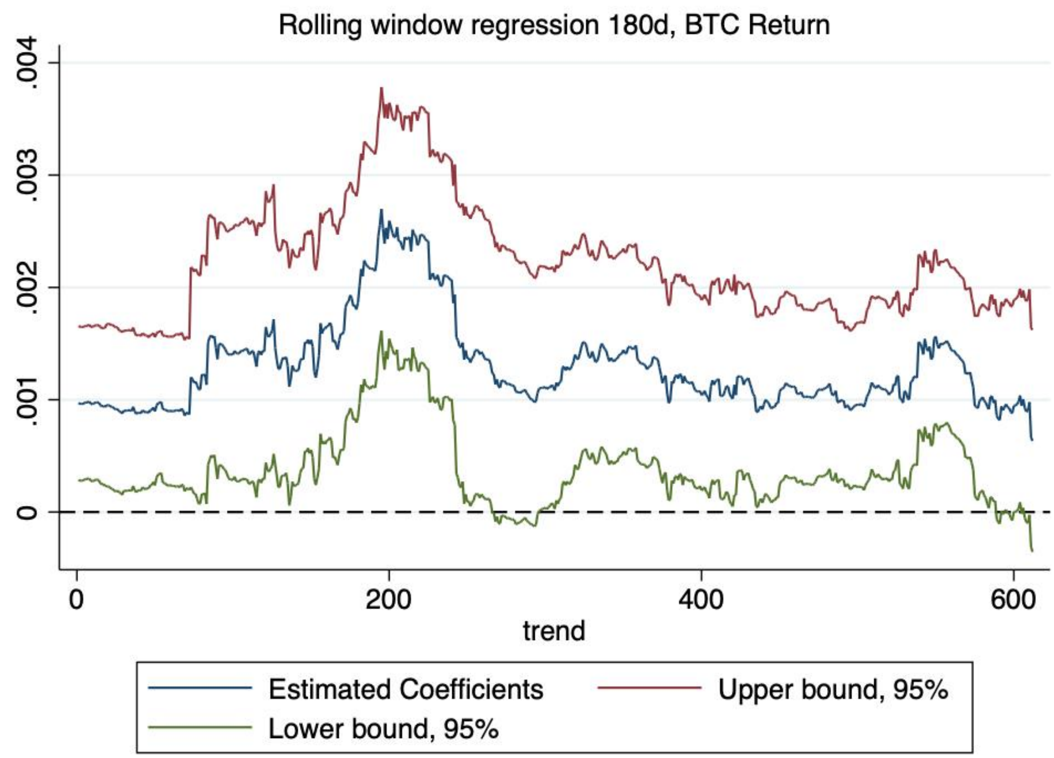
	Twitter Mean Compound Score $_{BTC, t}$	Negative tweets in % of total $_{BTC, t}$	Positive tweets in % of total $_{BTC, t}$
BTC Return	0.000939***	-0.000250***	0.000055
Lag 1	0.001313***	-0.000312***	0.000086
Lag 2	0.0005356	-0.000149***	-0.000017
Lag 3	0.000895***	-0.000210***	0.000042
Lag 4	0.000874***	-0.000167***	0.000080
Lag 5	0.000902***	-0.000049	0.000204**
Lag 6	0.000994***	-0.000051	0.000252
Lag 7	0.000744***	-0.000138***	0.000075
Lag 8	0.000417	-0.000073	0.000080
S&P 500$_t$	0.000878	-0.000144	0.000108
US 90d$_t$	-0.0000007**	5.15E-08	-0.0000002*
WOCU$_t$	-0.008172	0.000272	-0.002013
Observations	535	535	535
R^2	0.152	0.192	0.042
Adjusted R^2	0.132	0.174	0.020

*Notes: Parameters of linear regression, with standard deviation in parenthesis. *10%, **5% and ***1% significance level.*

The findings are consistent with the results we obtained in Tables 5.1.1 and 5.1.2, indicating that they are reliable. Only the lags with a higher significance level than 1% disappear for both mean and negative tweets. The regression for the proportion of positive tweets confirms our beliefs that the significant values result from randomness. Furthermore, we find that US 3 Month Treasury Bill is significant for the mean compound regression with a negative coefficient value, which is consistent with the fact that when interest rates decrease, people become more optimistic about Bitcoin as an investment and vice versa. We also observe that the adjusted R^2 for the negative proportion regression improves by two percentage points while the mean compound adjusted R^2 remains almost the same. Lastly, we find no significant relationship between sentiment and S&P 500 or WOCU.

In addition, the model's stability over time is assessed using a rolling window regression. We used a fixed size of 180 days when running the rolling window regression on Equation 4.2.1. Figure 5.3.1 plots β_1 , the Bitcoin return at time t, where the first estimated coefficient is from January 1, 2020, to June 28, 2021, and the last estimated coefficient is from September 3, 2021, to March 1, 2022. The corresponding figures for lag one to three is found in the Appendix.

Figure 5.3.1 Rolling Window Regression of the first coefficient in Equation 4.2.1 with a fixed number of 180 days.



A 5 percent significance level is used to compute the upper and lower bound of the confidence level. If the dotted line is between the lower and upper bound, the estimated coefficient is not statistically significantly different from zero. The graph shows that the estimated coefficients significantly differ from zero during most of the period, only with a few minor exceptions. The beta coefficient's maximum value is from July 13, 2020, to January 1, 2021. The average value of the 612 estimated coefficients is 0.0013.

The corresponding test for lag one, seen in the Appendix, shows similar results. The estimated coefficient is statistically significant from zero for most of the period and peaks from mid-2020 to the start of 2021. On the other hand, lag number two and three are in extensive periods not statistically significantly different from zero. The insignificance of lag number two is consistent with our findings when introducing additional variables in Table 5.3.1. Bitcoin returns from two or more days ago are not stable predictors of the sentiment score on Twitter.

We discussed in Section 5.2 that conservative people are more rational and better informed than the average analyst. Furthermore, Bonaparte et al. (2017) find evidence that when an investors' preferred party was in power, they increased their allocations to riskier assets. Our sample consists of times when Democrats and Republicans were in power in the United States. It will be interesting to observe if bitcoins predicting power on sentiment scores are different in those two periods. Notably, only 25 percent of Twitter users are Americans.

The results for when Republicans were in power, from January 1, 2020, to January 19, 2021, and when Democrats were in power, from January 20, 2021, to March 1, 2022, are seen in Table 5.3.2. We observe that there are significant differences between these two subsamples. All coefficients in the first subsample are significant, with an adjusted R^2 of 22.63 percent. In the second subsample, six coefficients are significant, but only one is on a significant level of 1 percent. The adjusted R^2 of 8.20 percent is relatively low. It is unreasonable to suppose that the political party in power causes the differences in the two subsamples, but it can back our discussion about political preference in Section 5.2.

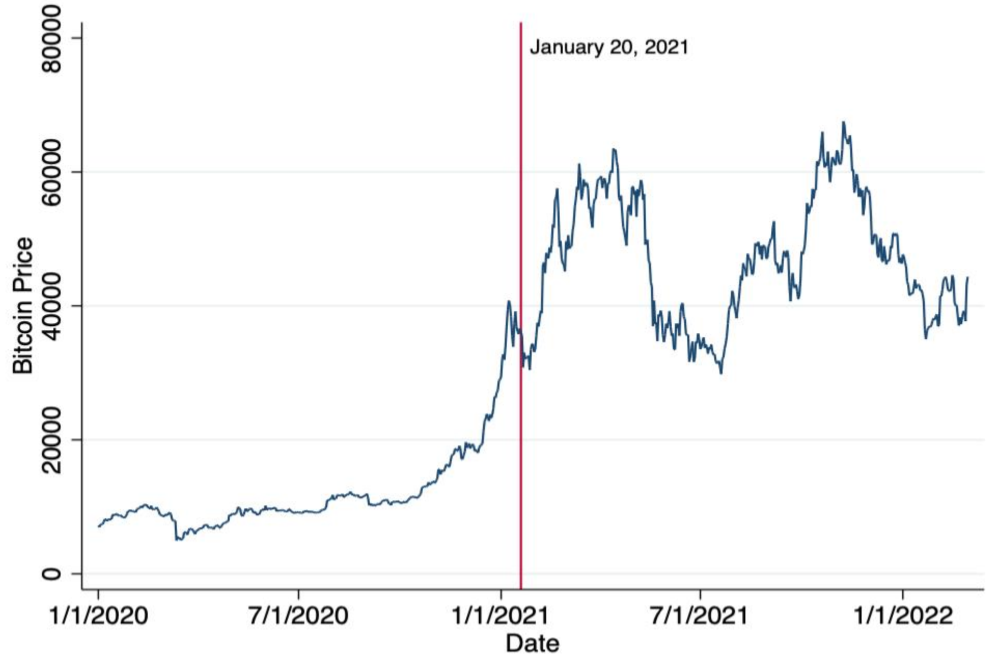
TABLE 5.3.2
LINEAR REGRESSION COEFFICIENTS OF TWO
SUBSAMPLES

	Republican 1/1/2020 – 19/1/2021	Democrats 20/1/2021 – 1/3/2022
BTC Return	0.001507***	0.000905**
	(0.00038)	(0.00036)
Lag 1	0.001502***	0.001134**
	(0.00036)	(0.00046)
Lag 2	0.000990***	0.000363
	(0.00036)	(0.00053)
Lag 3	0.001258***	0.000513
	(0.00038)	(0.00040)
Lag 4	0.001244***	0.001006***
	(0.00032)	(0.00038)
Lag 5	0.000920***	0.000738*
	(0.00033)	(0.00041)
Lag 6	0.000971**	0.001068***
	(0.00041)	(0.00036)
Lag 7	0.000839**	0.000865**
	(0.00036)	(0.00040)
Lag 8	0.000604*	0.000215
	(0.00032)	(0.00043)
Constant	0.093699	0.1034994
	(0.00150)	(0.00191)
Observations	377	398
R^2	0.2448	0.1028
Adjusted R^2	0.2263	0.0820
F-statistics	4.99*** (df: 9, 367)	4.14***(df: 9, 388)

Notes: Parameters of linear regression, with standard deviation in parenthesis.
**10%, **5% and ***1% significance level.*

If we compare the price movements of Bitcoin across the two periods, as seen in Figure 5.3.2, we can observe a significant difference. During the era when Republicans were in control, the price of Bitcoin increased significantly. In comparison, the price went more up and down in the second period. This might imply that there was more excitement around Bitcoin since the price achieved its all-time high, rather than who was in control.

Figure 5.3.2 Bitcoin Price Movement from January 1, 2020, to March 1, 2022



6 Conclusion

This thesis explores if and to what degree returns predict the sentiment scores on Twitter and Reddit. In our analysis, we have used a large and unique dataset containing tweets and posts on Reddit mentioning Bitcoin published from January 1, 2020, to March 1, 2022.

We use simple linear regression to analyze our hypothesis to predict the sentiment score at time t using Bitcoin returns with lags up to 8 days. We find a statistically significant relationship between Bitcoin returns, up to 7 lags, and the average sentiment score on Twitter. These findings show that a decrease in Bitcoin return predicts more negative tweets from individuals on Twitter and vice versa. However, by observing the percent of negative and positive tweets alone, will Bitcoin returns have significantly more impact on negative tweets than positive. This can indicate that Bitcoin returns have greater predicting power on how pessimistic individuals are on Twitter. A decrease either makes users go silent, or a small portion becomes positive. Regardless of whether the price of Bitcoin falls or rises, Bitcoin enthusiasts remain optimistic.

In addition, we find disparities in predicting power within the sample. From 2020 to the start of 2021, we observe a greater R^2 than in the remaining period. It is not clear why these disparities exist, but they might be related to the persistent price increase in the first part of the sample and the highs and lows in the second period. However, both periods' returns only explain some of the average sentiment scores. In the rolling window robustness check, we see that only the return on the same day and yesterday returns are stable predictors of the mean sentiment score.

Furthermore, we see apparent differences between Twitter and Reddit. It appears that returns do not affect the average sentiment score on Reddit, which might be due to different characteristics. We observe differences in age, gender, nationality, annual household income, and political preferences, all of which could indicate different risk preferences. However, these differences are for the average Twitter and Reddit user, not necessarily the differences between people who tweet about Bitcoin and post on r/Bitcoin. It is also possible that the disparities we detect are due to the weaknesses of our study, which we will discuss in the next section.

Previous research has established a relationship between sentiment expressed on social media and cryptocurrency prices. However, we believe this influence is diminished due to a larger share of institutional investors and because Bitcoin has become more known. The more large institution and companies engage in cryptocurrency, the more difficult it will be for individuals to influence the market. Despite this, our research suggests that returns impact Twitter sentiment scores. If sentiment scores affect returns, their relationship can be explained as a continuous cycle. It might explain why Bitcoin could become so large.

6.1 Limitations and Future Studies

Our constructed variables and analysis methodologies have limitations that propose future research. We discovered that VADER had trouble analyzing some non-English posts when we randomly sampled our data sample. This did not apply to all non-English languages, but a few posts received a neutral compound score.

Both Twitter and Reddit can contain pictures, GIFs, and other non-lexical items, which challenges the process of assigning the appropriate sentiment score to each post. It is especially problematic for Reddit, which contains many pictures with text. The text can differ in terms of sentiment from the text on the picture. The lexical items are often short and neutral, whereas the other content is either positive or negative. An example of this is given in the Appendix. It is unclear how to solve this problem, but it is likely that it will impact our regression output.

Examining additional periods is a natural extension of the analysis undertaken in this paper, especially when we see such large differences in predicting power between subsamples. Furthermore, it would be interesting to see if political parties in power influence the model or if the significant disparities between 2020 and 2021-2022 are merely a coincidence. In addition, it would be interesting to apply this model to different cryptocurrencies and other social media platforms such as TikTok.

Bibliography

Ajaz, T. & Kumar, A. S. (2019). Co-Movement In Crypto-Currency Markets: Evidence From Wavelet Analysis. *Financial Innovation, 5*(33). https://doi.org/10.1186/s40854-019-0143-3

Al-Jarrah, I. M. W., Al-Yahyaee, K. H., Kang, S. H., Rehman, M U. & Mensi, W. (2019). Time Frequency Analysis of the Commonalities Between Bitcoin and Major Cryptocurrencies: Portfolio Risk Management Implications. T*he North American Journal of Economics and Finance, 48,* 283-294. https://doi.org/10.1016/j.najef.2019.02.013

Al-Yahyaee, K., Mensi, W. & Yoon, S-M. (2018). Efficiency, multifractality, and the long-memory property of the Bitcoin market: A Comparative analysis with stock, currency and gold markets. *Finance Research Letters*, *27*, 228-234. https://doi.org/10.1016/j.frl.2018.03.017

Amboala, T., Gazali, H. M. & Ismail, C. M. (2018). Exploring the Intention to Invest in Cryptocurrency: The Case of Bitcoin. *International Conference on Information and Communication Technology for the Muslim World (ICT4M).* https://doi.org/10.1109/ICT4M.2018.00021

Auer, R. & Claessens, S. (2018). *Regulating Cryptocurrencies: Assassin Market Reactions*. (BIS Quarterly Review September 2018). https://www.bis.org/publ/qtrpdf/r_qt1809f.pdf

Bariviera, A. F. (2017). The inefficiency of Bitcoin revisited: A dynamic approach. *Economic Letters*, *161*, 1-4. https://doi.org/10.1016/j.econlet.2017.09.013

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., Blackburn, J. (2020). The Pushshift Reddit Dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*(1), 830-839. https://ojs.aaai.org/index.php/ICWSM/article/view/7347

BIS (2021). *Distrust or speculation? The socioeconomic drivers of US cryptocurrency investments.* (Bank for International Settlements, Monetary and Economic Department, Working Paper No. 951). https://www.bis.org/publ/work951.pdf

Bloomenthal, A., Anderson, S. & Velasquez, V. (2021). What Determines the Price of 1 Bitcoin? *In Investopedia.* https://www.investopedia.com/tech/what-determines-value-1-bitcoin/

Bjørgan, H. C., Nakken, K. G. & Dukefoss, E. B. (2021). Predicting Cryptocurrency Returns Using Market, Social Media, Search Volume and Blockchain Data. [Master thesis, Norwegian University of Science and Technology]. NTNU Open. https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2787829/no.ntnu%3Ainspera%3A85279552%3A85282676.pdf?sequence=1

Brooks, C. (2019). *Introductory Econometrics for Finance* (4th ed.). Cambridge University Press. https://doi.org/10.1017/9781108524872

Britzolakis, A., Kondylakis, H. & Papadakis, N. (2020). A Review on Lexicon Based and Machine Learning Political Sentiment Analysis Using Tweets. *International Journal of Semantic Computing, 14*(4), 517-563. https://doi.org/10.1142/S1793351X20300010

Brunner, N. (n.d.). *Data: At What Age Do Students Graduate from College.*Salarship. https://salarship.com/data/average-age-college-graduates/

Bonaparte, Y., Kumar, A. & Page, J. K. (2017). Political Climate, Optimism, and Investment Decisions. *Journal of Financial Markets, 34*, 69-94. https://doi.org/10.1016/j.finmar.2017.05.002

Bonta, V., Kumaresh, N. & Janardhan, N. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. *Asain Journal of Computer Science and Technology, 8(S2), 1-6.* 10.51983/ajcst-2019.8.S2.2037

Borg, A. & Boldt, M. (2020). Using VADER sentiment and SVM for predicting customer response sentiment. *Expert Systems with Application, 162,* Article 113746. https://doi.org/10.1016/j.eswa.2020.113746

Bouri, E., Shahzad, S. J. H., Roubaud, D., Kristoufek, L. & Luecy, B. (2020). Bitcoin, gold, and commodities as safe havens for stocks: New insight through wavelet analysis. *The Quarterly Review of Economics and Finance, 77*, 156-164. https://doi.org/10.1016/j.qref.2020.03.004

Chainalysis (2021, January 19). Crypto Crime Summarized: Scams and Darknet Markets Dominated 2020 by Revenue, But Ransomware Is the Bigger Story. *Chainalysis.*

https://blog.chainalysis.com/reports/2021-crypto-crime-report-intro-ransomware-scams-darknet-markets/

Choudhury, S. R. (2021, January 4). Bitcoin is breaking records because bigger investors are buying now, says PwC. *CNBC.* https://www.cnbc.com/2021/01/04/bitcoin-btc-rally-partly-driven-by-more-institutional-investors-pwc-says.html

Ciaian, P., Rajcaniova, M. & Kancs, D. (2015). The Economics of Bitcoin Price Formation. *Applied Economics 48(19)*, 1799-1815. https://doi.org/10.1080/00036846.2015.1109038

Coinbase (2022). *Bitcoin price.* Coinbase. https://www.coinbase.com/price/bitcoin

Corbet, S., McHugh. G. & Meegan, A. (2017). The influence of central bank monetary policy announcement on cryptocurrency return volatility. *Investment Management and Financial Innovations, 14(*4), 60-72. http://dx.doi.org/10.21511/imfi.14(4).2017.07

Danling, J., Kumar, A. & Law, K. K. F. (2016). Political contributions and analyst behavior. *Review of Accounting Studies*, *21*, 37-88. https://doi.org/10.1007/s11142-015-9344-9

Dantes, D. (March 15, 2021). *Bitcoin Q1 Retail Flow Exceeding Institutional Investment*. Yahoo Finance. https://finance.yahoo.com/news/bitcoin-retail-flows-track-exceed-174119640.html

Dowling, M. (2022). Is non-fungible token pricing driven by cryptocurrencies? *Finance Research Letters, 44.* https://doi.org/10.1016/j.frl.2021.102097

Ebargir, S. & Yang, J. (2019), Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment. *IAENG Transactions on Engineering Science, Special Issue for the International Association of Engineers Conference 2019,* 63-80. https://doi.org/10.1142/9789811215094_0005

Edwards, J., Mansa, J. & Kvilhaug, S. (2022). Bitcoin's Price History. In *Investopedia.* https://www.investopedia.com/articles/forex/121815/bitcoins-price-history.asp

European Central Bank. (2021). *What is Bitcoin?* ECB. https://www.ecb.europa.eu/ecb/educational/explainers/tell-me/html/ what-is-bitcoin.en.html

Evite, P. M. (2018). *Efficient Market Hypothesis: Evidence from the Bitcoin Marke (2010-2017)*. [Bachelor thesis, University of the Philippines Los Banos]. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3830961

Fama, E. (1969). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25*(2), 383-417. https://doi.org/10.2307/2325486

Foley, S., Karlsen, J. R. & Putnins, T. J. (2019). Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed Through Cryptocurrencies? *The Review of Financial Studies, 44*(5), 1798-1853. https://doi.org/10.1093/rfs/hhz015

Glaser, F., Haferkorn, M., Siering, M., Zimmermann, K. & Weber, M. C. (2014). Bitcoin – Asset or Currency? Revealing Users' Hidden Intentions. *Twenty Second European Conference on Information Systems, Tel Aviv 2014*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2425247

Global Financial Data. (2022). S&P 500. [Historical Dataset]. https://globalfinancialdata.com/

Global Financial Data. (2022). 3 Month Treasury Bill Rate. [Historical Dataset]. https://globalfinancialdata.com/

Gonzalesz, O. & Priest, D. (2021, March 17). *Roobinhood backlash: What you should know about the GameStop controversy.* CNET. https://www.cnet.com/personal-finance/investing/robinhood-backlash-what you-should-know-about-the-gamestop-stock-controversy/

Hajric, V. (November 11, 2021). *Need an Inflation Hedge? Bitcoin Has Delivered 99.996% Deflation*. Bloomberg. https://www.bloomberg.com/news/articles/2021-11-11/need-an-inflation-hedge-bitcoin-has-delivered-99-996-deflation

Hayes, A., Brown, J. R. & Kazel, M. (March 5, 2022). What Happens to Bitcoin After All 21 Million are Mined? *In Investopedia.*

https://www.investopedia.com/tech/what-happens-bitcoin
-after-21-million-mined/

Hayes, A. S. (2017). Cryptocurrency Value Formation: An Empirical Study to a Cost of Production Model for Valuing Bitcoin. *Telematics and Informatics 34*(7), 1308-1321.
https://doi.org/10.1016/j.tele.2016.05.005

Hayes, A., Westfall, P. & Kvilhaug, S. (2022). Heteroskedasticity. In *Investopedia.*
https://www.investopedia.com/terms/h/heteroskedasticity.asp

Hens, T., Rieger, M. O. & Wang, M. (2011). Prospect Theory Around the World. *NHH Dept. of Finance & Management Science Discussion Paper No.2011/19*.
https://dx.doi.org/10.2139/ssrn.1957606

Hicks, C. & Likos, P. (2022, February 4). *The History of Bitcoin, the First Cryptocurrency.* U.S.News Money.
https://money.usnews.com/investing/articles/the-history-of-bitcoin

Hinchliffe, R. (2021, June 16). *What drives the value of crypto?* Raconteur.
https://www.raconteur.net/finance/investing/what-drives-value
cryptocurrencies/

Hutto, C. (2021). *VaderSentiment.* Github.
https://github.com/cjhutto/vaderSentiment/blob/master/README.rst

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, 8*(1), 216-225.
https://ojs.aaai.org/index.php/ICWSM/article/view/14550

Iyer, T. (2022). Cryptic Connections: Spillovers between Crypto and Equity Markets. *Global Financial Stability Notes, 1*.
https://www.imf.org/en/Publications/global-financial-stability-notes
/Issues/2022/01/10/Cryptic-Connections-511776

Jareño, F., Gonzáles, M., Tolentino, M. & Sierra. K. (2020). Bitcoin and gold price returns: A quantile regression and NARDL analysis. *Resources Policy, 67*, Article 101666.
https://doi.org/10.1016/j.resourpol.2020.101666

Jiang, Y., Nie, H. & Ruan, W. (2018). Time-varying long-term memory in bitcoin market. *Finance Research Letters, 25*, 280–284.
https://doi.org/10.1016/j.frl.2017.12.009

Jianakoplos, N. A. & Bernasek, A. (2007). Are women more risk averse? *Economic Impurity*,*36*(4), 620-630. https://doi.org/10.1111/j.1465-7295.1998.tb01740.x

Kannadhasan, M. (2015). Retail Investors' Financial Risk Tolerance and Their Risk Taking Behaviour: The Role of Demographics as Differentiating and Classifying Factors. *IIMBManagement Review, 27*(3), 175-184. https://doi.org/10.1016/j.iimb.2015.06.004

Katsiampa, P. (2019). Volatility Co-Movement Between Bitcoin and Ether. *Finance ResearchLetters, 30*, 221-227. https://doi.org/10.1016/j.frl.2018.10.005

Kim, Y. B., Kim, J. G., Kim, W., Im, H. I., Kim, T. H., Kang, S. J. & Kim, C. H. (2016). Predicting Fluctuations in Cryptocurrency Transaction Based on User Comments and Replies. *PLoS ONE 11*(8). https://doi.org/10.1371/journal.pone.0161197

Kim, S. & Kim, D. (2014). Investor Sentiment From Internet Message Postings and the Predictability of Stock Returns. *Journal of Economic Behavior & Organization*, *107*(B), 708-729. https://doi.org/10.1016/j.jebo.2014.04.015

Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports, 3*(1), Article 3415. 10.1038/srep03415

Konermann, P. (2021). *GRA 6547 Research Methodology in Finance.* [PowerPoint presentation]. Itslearning. https://bi.itslearning.com/

Koutmos, D. Market risk and Bitcoin returns. *Annals of Operations Research, 294*, 453-477. https://doi.org/10.1007/s10479-019-03255-6

Leech, O. (2021, October 26). *Want to Buy Dogecoin? Read This First*. CoinDesk. https://www.coindesk.com/learn/want-to-buy-dogecoin-read-this-first/

Li, Y. (2021, January 27). GameStop mania explained: How the Reddit retail trading crowd ran over Wall Street pros. *CNBC.* https://www.cnbc.com/2021/01/27/gamestop-mania-explained-how-the-reddit-retail-trading-crowd-ran-over-wall-street-pros.html

Liu, Y. & Tsyvinski, A. (2018). Risks and Returns of Cryptocurrency. https://dx.doi.org/10.2139/ssrn.3226952

Luo, X., Zhang, J., & Duan, W. (2013). Social Media and Firm Equity Value. *Information Systems Research*, *24*(1), 146–163. https://doi.org/10.1287/isre.1120.0462

Mai, F., Shan, Z., Bai, Q., Wang X. & Chiang, R. (2018). How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis. *Journal of Management Information Systems*, *35*(1), 19-52 https://doi.org/10.1080/07421222.2018.1440774

Nadarajah, S. & Chu, J. (2017). On the inefficiency of Bitcoin. *Economic Letters, 150*, 6-9. https://doi.org/10.1016/j.econlet.2016.10.033.

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf

Nietzel, M. T. (2021, February 22). New from U.S. Census Bureau: Number of Americans With A Bachelor’s Degree Continues To Grow. *Forbes*. https://www.forbes.com/sites/michaeltnietzel/2021/02/22/new-from-us-census-bureau-number-of-americans-with-a-bachelors-degree-continues-to-grow/?sh=5d95a0d87bbc

Ostroff, C. (2022, January 6). Bitcoin, Ether Prices Fall Along With Tech Stocks. *The Wall Street Journal.* https://www.wsj.com/articles/bitcoin-ether-prices-fall-along-with-tech-stocks-11641471483?reflink=e2twmkts

Pew Research Center. (2016, February 25). *Reddit news users more likely to be male, young and digital in their news preference.* Pew Research. https://www.pewresearch.org/journalism/2016/02/25/reddit-news-users-more likely-to-be-male-young-and-digital-in-their-news-preferences/

Pew Research Center. (2018, September 13). *Republicans, Democrats See Opposing Party as More ideological Than Their Own.* Pew Research. https://www.pewresearch.org/politics/2018/09/13/republicans -democrats-see-opposing-party-as-more-ideological-than-their-own/

Pano, T. & Kashef, R. (2020). A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweet during the Era of COVID-19. *Big Data Cogn. Comput.*, *4*(4), 33.

https://doi.org/10.3390/bdcc4040033

Qin, A. & Livni, E. (2021, September 24). China Cracks Down Harder on Cryptocurrency with New Ban. *The New York Times. https://www.nytimes.com/2021/09/24/business/china-cryptocurrency-bitcoin.html*

Reddit. (2022). *Wallstreetbets.* Reddit. https://www.reddit.com/r/wallstreetbets/

Reddit. (2022). *Press.* Reddit. https://www.redditinc.com/press

Regnier, P. (2021, February 4). Stonks Are Bonkers, and Other Lessons from the Reddit Rebellion. *Bloomberg.* https://www.bloomberg.com/news/features/2021-02-04/gamestop-gme-how-wallstreetbets-and-robinhood-created-bonkers-stock-market

Røkke, K. I. (2021). *Shareholder Letter.* Seetee. https://www.seetee.io/static/shareholder_letter6ae7e85717c28831bf1c0eca1d32722.pdf

Saad, S. E. & Yang. J. (2019). Twitter Sentiment Analysis Based on Ordinal Regression. *IEEE Access* , *7*(1), 163677-163685. 10.1109/ACCESS.2019.2952127

Saleh, M. (2018). *Cryptonomics: Investment behaviour in the cryptocurrency market* (Working paper 2018). https://www.academia.edu/37876981/Work_in_Progress_Cryptonomics_Investment_behaviour_in_the_cryptocurrency_market

SAS. (n.d.). *Natural Language Processing (NLP).* SAS. https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html

Sawhney, K. (2021, January 16). Don't Let Crypto Mayhem Spook Retail Investors. *Fortune.* https://fortune.com/2022/01/26/dont-let-crypto-mayhem-spook-retail-investors-sell-off-news-bitcoin-investing/

Seo, E-J. & Park, J-W. (2018). A study on the effects of social media marketing activities on brand equity and customer response in the airline industry. *Journal of Air Transport Management, 66*, 36-41.

https://doi.org/10.1016/j.jairtraman.2017.09.014

Sharma, R., Clemon, D. & Rathburn, P. (2021). Non-Fungible Token (NFT) Definition. *In Investopedia*. https://www.investopedia.com/non-fungible-tokens-nft-5115211

Shanaev, S., Sharma, S., Ghimire, B. & Shuraeva, A. (2020). Taming the blockchain beast? Regulatory Implications for the Cryptocurrency Market. *Research in International Business and Finance, 51*. https://doi.org/10.1016/j.ribaf.2019.101080.

Statistia. (2022). *Bitcoin market dominance - its market cap relative to the market cap of all other cryptocurrencies in the world - from April 2013 up until May 16, 2022.* [Statistics]. https://www.statista.com/statistics/1269669/bitcoin-dominance-historical-development/

Statistia. (2022). *Twitter: distribution of global audience 2021, by age group.* [Statistics]. https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/

Statistia. (2022). *Global Reddit users distribution by gender 2022.* [Statistics]. https://www.statista.com/statistics/1255182/distribution-of-users -on-reddit-worldwide-gender/

Statistia. (2022). *Global Twitter user distribution by gender 2022.* [Statistics]. https://www.statista.com/statistics/828092/distribution-of-users -on-twitter-worldwide-gender/

Statistia. (2022). *Regional distribution of desktop traffic to Twitter.com as of February 2022, by country.* [Statistics]. https://www.statista.com/statistics/261402/distribution-of-twitter-traffic-by-country/

Statistia. (2022). *Regional distribution of desktop traffic to Reddit.com as of February 2022, by country.* [Statistics]. https://www.statista.com/statistics/325144/reddit-global-active-user-distribution/

Snscrape (2022). *Github.* https://github.com/JustAnotherArchivist/snscrape

Subramaniam, S. & Chakraborty, M. (2019). Investor Attention and Cryptocurrency Returns: Evidence from Quantile Causality Approach. *Journal of Behavioral Finance, 21*(1), 103-115. https://doi.org/10.1080/15427560.2019.1629587

Tiwari, A. K., Jana, R. K., Das, D. & Roubaud, D. (2018). Informational efficiency of bitcoin - an extension. *Economics Letters, 163*, 106–109. https://doi.org/10.1016/j.econlet.2017.12.006

Tiwari, A. K., Raheem. I. D., Kang, S. H. (2019). Time-varying dynamics conditional correlation between stock and cryptocurrency markets using the copula ADCC-EGARCH model.*Physica A: Statistical Mechanics and its Applications, 535*, 122295. https://doi.org/10.1016/j.physa.2019.122295

Todorov, G. (2022, March 14). 70+ Important Reddit Statistics 2020. *Thrive my way.* https://thrivemyway.com/reddit-statistics/

Tonkin, E. L. (2016). A Day at work (with Text): A Brief Introduction. *Working with Text*, 23-60. https://doi.org/10.1016/B978-1-84334-749-1.00002-0

Turmarkin, R. & Whitelaw, R. F. (2001). News or Noise? Internet Postings and Stock Prices. *Financial Analysts Journal, 57*(3), 41-51. https://www.jstor.org/stable/4480315?seq=1

Twitter (2022). *Fiscal Year Annual Report 2021.* https://s22.q4cdn.com/826641620/files/doc_financials/2021/ar/FiscalYR2021_Twitter_Annual_-Report.pdf

Urquhart, A. (2016). The inefficiency of Bitcoin. *Economic Letters, 148,* 80-82. https://doi.org/10.1016/j.econlet.2016.09.01

Urquhart, A. (2018). What Causes the Attention of Bitcoin? *Center of Digital Finance 166,*40-44. https://doi.org/10.1016/j.econlet.2018.02.017

Voell, Z. (2021, February 25). *Coinbase Institutional, Retail Trading Volume Grew at Equal Rates in 2020.* CoinDesk. https://www.coindesk.com/markets/2021/02/25/coinbase-institutional-retail-trading-volume-grew-at-equal-rates-in-2020/

Weisenthal, J. (May 4, 2021). Two Ways That Dogecoin Is the GameStop of Cryptos. *Bloomberg.* https://www.bloomberg.com/news/articles/2021-05-04/two-ways-that-dogecoin-is-the-gamestop-of-cryptos

WOCU. (n.d.). *WOCU World Currency Unit*. WOCU. http://www.wocu.com/

Wolfson, R. (April 11, 2022). *Grassroots initiatives are bringing Bitcoin education to communities across America.* Cointelegraph. https://cointelegraph.com/news/grassroots-initiatives-are-bringing-bitcoin-education-to-communities-across-america

Wolk, K. (2019). Advanced social media sentiment analysis for short-term cryptocurrency price prediction. *Expert Systems 37(2).* https://doi-org.ezproxy.library.bi.no/10.1111/exsy.12493

Wolla, S. A. (2018). *Bitcoin: Money or Financial Investment?* Page One Economics. https://files.stlouisfed.org/files/htdocs/publications/page1-econ/2018/03/01/bitcoin-money-or-financial-investment_SE.pdf

Wilhelm, A. (2022, January 6). *How will the crypto selloff impact the NFT market?* TechCrunch. https://techcrunch.com/2022/01/06/how-will-the-crypto-selloff-impact-the-nft-market/

Wojcik, S. & Hughes, A. (2019, April 24). Sizing up Twitter Users. *Pew Research Center.* https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/04/twitter_opinions_4_18_final_clean.pdf

Yahoo Finance. (2022). Bitcoin. Ethereum. Monero. Dogecoin. Shiba Inu. [Historical Data]. Yahoo Finance. https://finance.yahoo.com/

Yahoo Finance. (2022, June 3). *GameStop: Investors Have No Way to Assess the Company's Intrinsic Value, Says Analyst.* Yahoo Finance. https://finance.yahoo.com/news/gamestop-investors-no-way-assess 225803895.html

Yermack, D. (2015). Is Bitcoin a Real Currency? An Economic Appraisal. *Handbook of Digital Currency*, 31-43.

https://doi.org/10.1016/B978-0-12-802117-0.00002-3

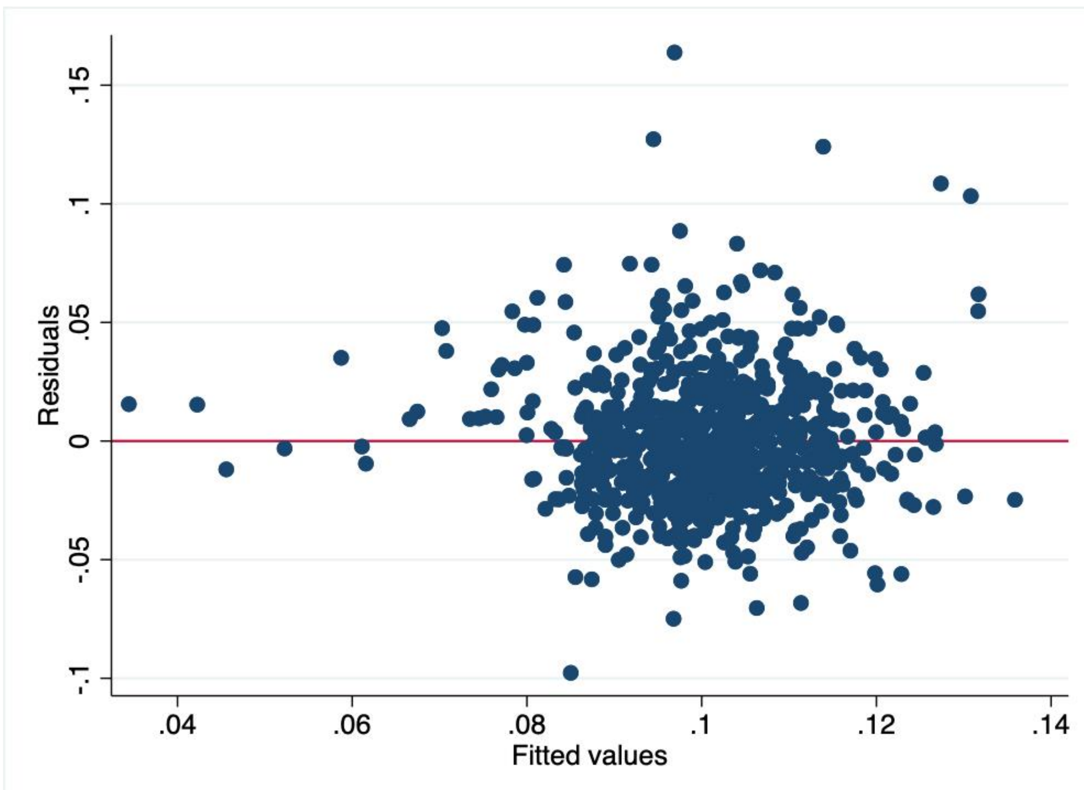
Zeng, T., Yang, M. & Shen, Y. (2020). Fancy Bitcoin and conventional financial assets: Measuring market integration based on connectedness networks. *Economic Modelling*, *90*, 209-220. https://doi.org/10.1016/j.econmod.2020.05.003.

Appendices

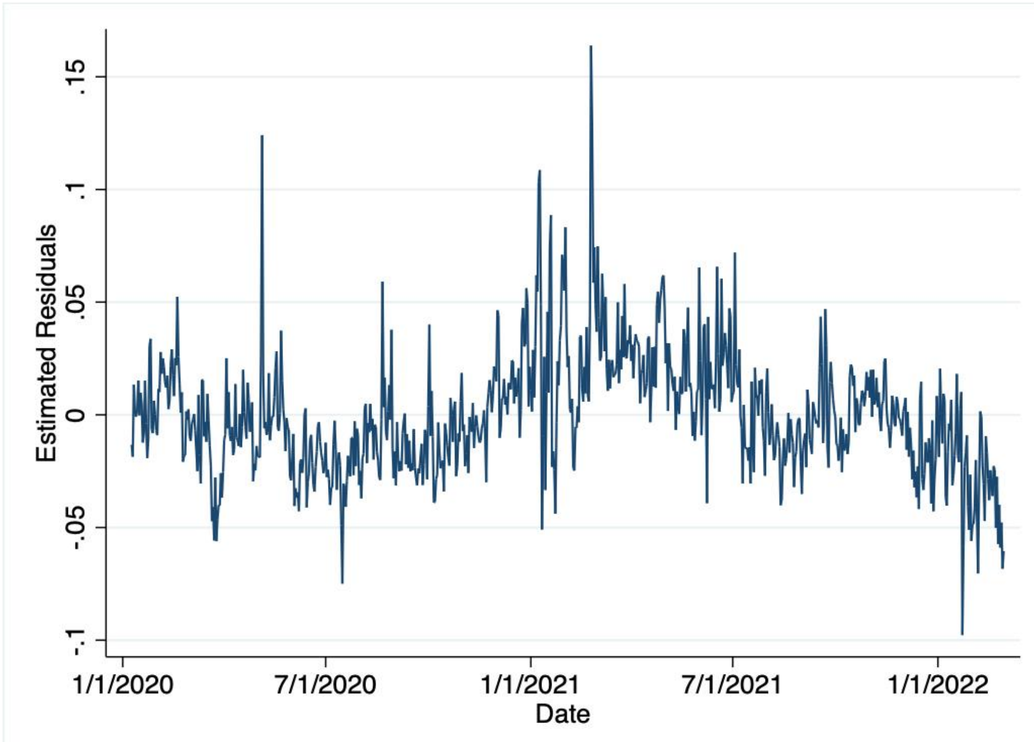
Table 1: Correlation between estimated residuals and independent variables.

BTC Return lag	Correlation
BTC Return	0.00
Lag 1	0.00
Lag 2	0.00
Lag 3	0.00
Lag 4	0.00
Lag 5	0.00
Lag 6	0.00
Lag 7	0.00
Lag 8	0.00

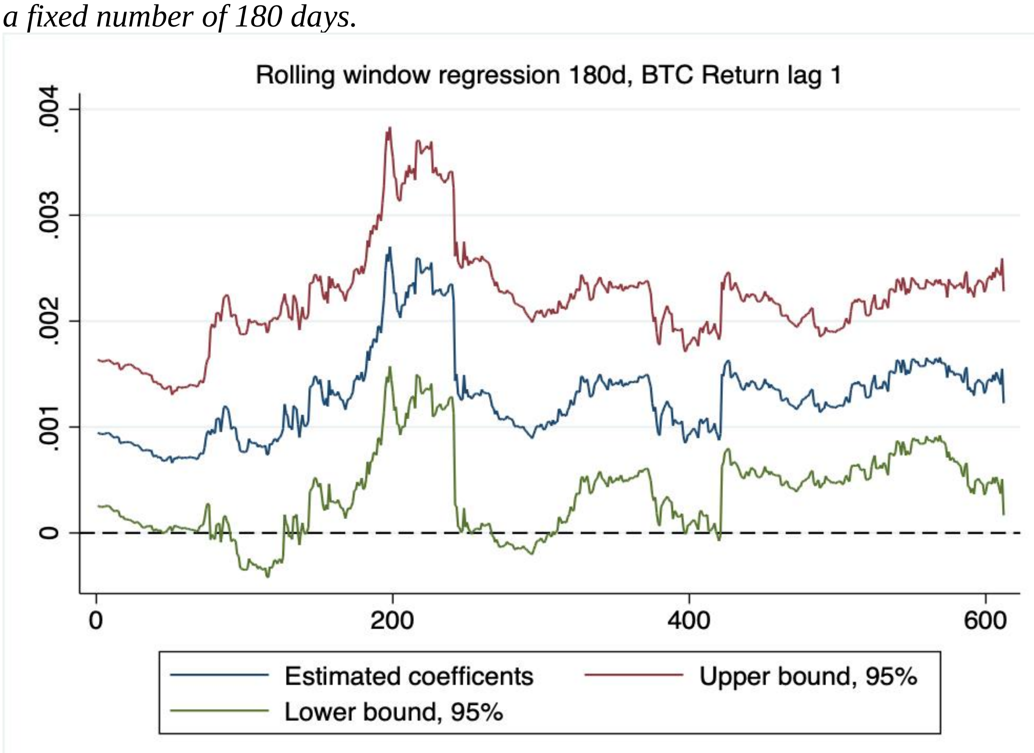
Graph 1: Residual-versus-fitted plot of Equation 4.2.1.



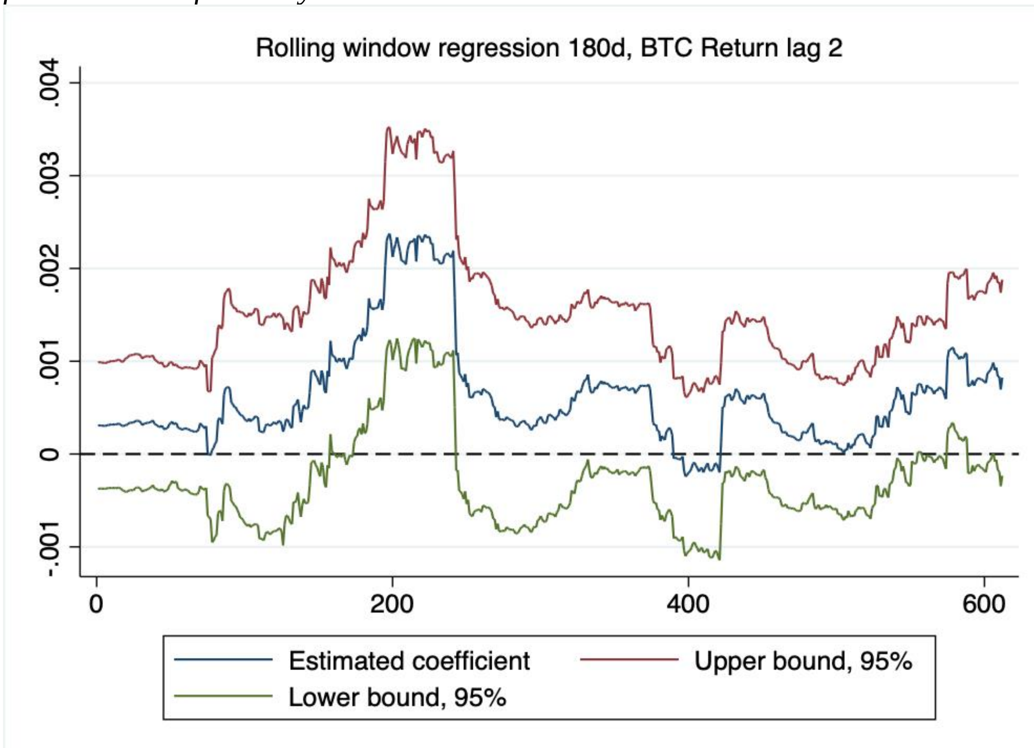
Graph 2: Estimated residuals from Equation 4.2.1.



Graph 3: Rolling Window Regression of the second coefficient in Equation 4.2.1 with a fixed number of 180 days.



Graph 4: Rolling Window Regression of the third coefficient in Equation 4.2.1 with a fixed number of 180 days.



Graph 5: Rolling Window Regression of the fourth coefficient in Equation 4.2.1 with a fixed number of 180 days.

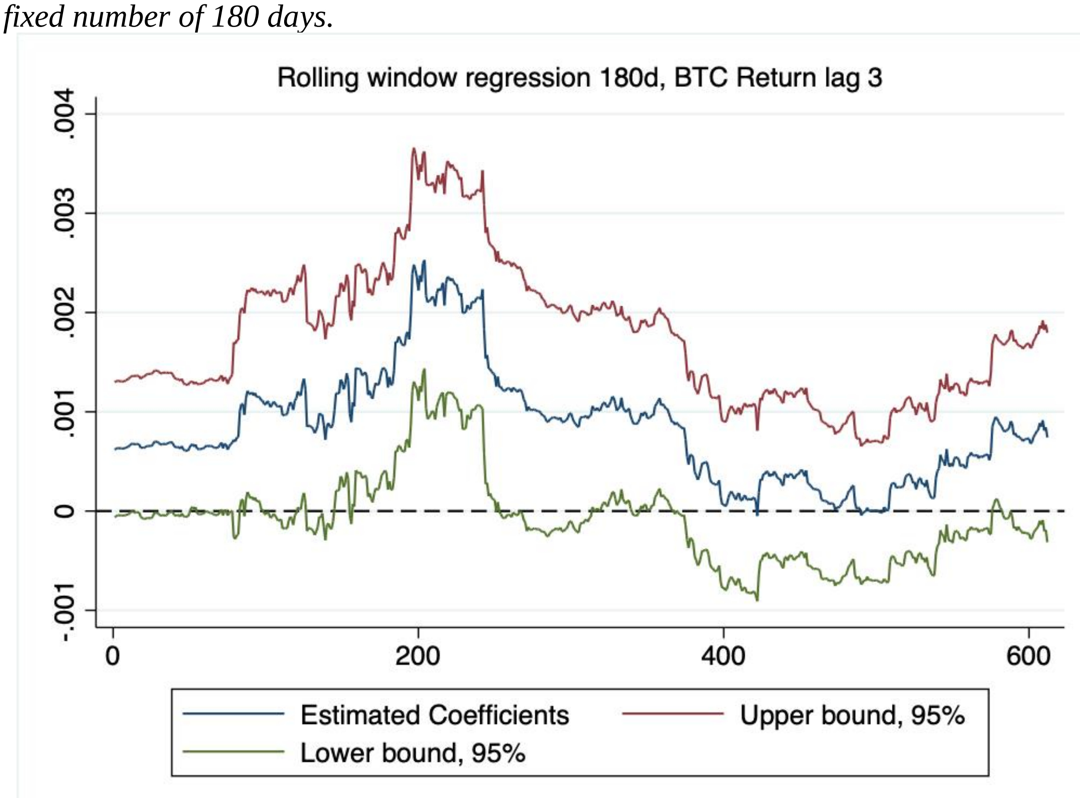


Figure 1: Example of Reddit post.