Whose opinions prevail on Bitcoin pricing?

Yang Yang

HANKEN SCHOOL OF ECONOMICS

Department of: Accounting and commercial law	**Type of work:** Thesis
Author: Yang Yang	**Date:** 15.1.2020
Title of thesis: Whose opinions prevail on Bitcoin pricing?	

Abstract: Bitcoin as a new asset class poses great challenges on existing asset pricing models. The neoclassical finance cannot explain the market anomalies around Bitcoin. Hence many researchers attempt to apply the findings of behavioural finance to Bitcoin pricing. News sentiment, whose impact on stock returns has been extensively studied, now also inspires a new strand of studies about its impact on Bitcoin pricing. This thesis collects news articles published by South China Morning Post and Financial Times from 2013 to 2019, in an attempt to compare the magnitude of sentiment impact from two continents, Asia and Europe, on Bitcoin price formation. The return of Bitcoin responds positively to both positive and negative news at a monthly frequency. Positive sentiment has almost twice the impact on Bitcoin monthly return than negative sentiment does. Both news sentiment and news volume have significant positive impact on Bitcoin prices at weekly frequencies. Investor attention proxied through news volume has considerably higher impact on Bitcoin prices than investor sentiment proxied through news sentiment. The continental differences of investor opinions between Asia and Europe are observed neither at a weekly nor at a monthly frequency. Cross-validation using the alternative attention measure, SVI, still fails to conclude the continental difference of investor sentiment.

Keywords: bitcoin pricing, news sentiment, investor attention, SVI, SCMP, FT, textual analysis

CONTENTS

APPENDICES

TABLES

FIGURES

1   INTRODUCTION

Bitcoin was conceived by a mysterious entity under the name Satoshi Nakamoto in 2008. This entity may be an individual or a group of people. But whoever is behind the design of Bitcoin appears to be not only very paranoid about privacy but also sufficiently frustrated by the prevalent fractional-reserve banking system. One may conjecture their motives from the two prime technologies of Bitcoin: the incorporation of public key infrastructure allowing secure authentication without revealing users' identities and the creation of blockchain technology which decentralizes currency issuance, resolves trust issues and intermediaries. In my opinion, the combination of these two technologies defines the fundamental value of Bitcoin.

However, the tremendous uncertainty and economic implications opened up by the new technologies impose great challenges on applying classical asset pricing theories and makes the price of Bitcoin extremely volatile. With a dollar value of $0.3 per Bitcoin in January 2011, the exchange rate skyrocketed over $19700 by the end of 2017. The high exchange rate volatility suggests that Bitcoin is not utilized as an alternative transaction system as intended by its creators, but rather as a speculative digital asset. Before a sensible pricing model is established, investors are forced to value Bitcoin on any information available, including social media, news articles, internet communities and even friends.

1.1   Problem

1.1.1   The hunting grounds of investor sentiment

Due to the speculative nature, sentiment has been considered as a significant driving factor for the Bitcoin prices. The impact of sentiment on stock markets has been extensively studied in the last decades (Barberis, et al., 1998) (Tetlock, 2007) (Baker & Wurgler, 2007) (Barber & Odean, 2008) (Fang & Peress, 2009) (Chousa, et al., 2016). The sentiment-based hypothesises now inspire a new strand of researches towards the pricing of cryptocurrencies. Abraham et al find Twitter sentiment cannot predict Bitcoin price (Abraham, et al., 2018). While Nasekin et al. find the sentiment index they constructed using StockTwits data significantly contribute to cryptocurrency's log return (Nasekin & Chen, 2019). Valencia et al. also observe the predictive power of sentiment on the direction of cryptocurrency market movement using data from cryptocompare.com (Valencia, et al., 2019). There are at least two fundamental issues in

quantifying the investor sentiment towards Bitcoin: first, where to aggregate people's sentiments towards Bitcoin; second, how to accurately measure these sentiments.

Social media platforms, e.g., Twitter (Zhang, et al., 2011), StockTwits (Nasekin & Chen, 2019), Facebook (Karabulut, 2012), have become natural places to capture public sentiment towards certain topics due to the large amount of users and convenient APIs for data queries. Moreover, open source packages specifically developed for analysing sentiment of social media contents are readily available and easy to use. But such internet-expressed sentiment is often noisier than news articles, since a large proportion of internet messages are posted by uninformed investors who might be susceptible to particular opinions and sentiments (Kearney & Liu, 2014).

Undoubtably, texts from social media platforms directly convey people's feelings since they are written by individual users (except the messages generated by computers). But what evokes or influences people's sentiment? According to the Agenda-setting theory developed by McCombs and Shaw, media has a great influence to its audience, "the media's agenda set the public's agenda" (McCombs & Reynolds, 2002). People's decisions are constantly influenced by the news they come across. Hence news from popular news agencies become ideal research candidates of investor sentiment. The news from trusted sources may contain valuable information worth being incorporated into asset pricing. On the other hand, news may be reflecting public opinions since news often succumb to public interest to attract readers. As observed by Ahern et al. newspapers are incentivized to publish sensational stories to grab attention and compete for readership (Ahern & Sosyura, 2013). Financial articles from major news agencies, e.g., Wall Street Journal (Tetlock, 2007) or Financial Times (Ferguson & Philip, 2015) (Kelly & Ahmad, 2018) (Smales & Lucey, 2019) have been reported to have significant influence on investor sentiment.

Investor attention (which is also categorized as sentiment by some researchers) is another critical factor investigated in this study. Attention is a scarce cognitive resource. Merton's theory predicts that attention could increase market valuations directly by alleviating informational friction that prevent investors from holding lesser-known assets (Merton, 1987). Barber and Odean find that retail investors are net buyers of stocks featured in Dow Jones News Service because of the limits on attention and short sales (Barber & Odean, 2008). Both theories predict increases in investor attention lead to increases in valuation and decreases in future returns.

According to McCombs et al., media attempts to establish a hierarchy of news prevalence. In other words, if the occurrence of certain news item is higher, the audience will regard this issue as more important (McCombs & Reynolds, 2002). Hence, the higher the news volume, the more attention investor will pay to a certain issue. Fang et al. used firm-specific news articles from several news agencies as a proxy for investor attention and found that stocks without media coverage in the prior month earn 3% higher annualized returns than stocks with above-average media coverage (Fang & Peress, 2009). According to Merton's theory, media coverage could make everyday investors aware of certain relatively obscure stocks.

Sentiment analysis is usually used to classify the polarity of a given text-whether the sentiment in the text is positive, negative or neutral. In general, there exists two approaches to measure sentiment, roughly divided into lexicon-based methods and machine-learning methods. Each method has its pros and cons (see section 2.4.1 for more detailed discussion). The accuracy of a sentiment analysis system is usually evaluated by comparing against human judgments. Classifying the polarity of texts typically achieves an accuracy of about 70% to 90%, depending on topical area (Karlgren, et al., 2012).

1.1.2   The different opinions from Asia and Europe

Despite the advancement of globalization and the diminishing isolation of domestic markets, the economic development in Asia is still very different from that in Europe in various aspects. Understanding this difference in the context of the increased financial linkage among countries from different continents adds to insights on Bitcoin pricing.

China and UK as the representative countries in their own continents are studied in this thesis. Relative to European countries, China has imposed more regulations on digital currencies. As summarised by Ponsford, financial institutions in China are banned from processing Bitcoin transactions and The People's Bank of China (PBOC) does not provide legal status nor recognise Bitcoin as a currency (Ponsford, 2015). Such radical rulings on Bitcoin exacerbated price fluctuations not only in China but also globally. Europe, on the other hand, is more concerned about how to tax cryptocurrencies while recognizing the legal status of Bitcoin.

The regulatory and economic differences between Asia and Europe likely lead to a different sentiment between Asian investors and European investors, which may be

captured by the news articles published by South China Morning Post (SCMP) and Financial Times (FT) respectively.

1.2   Purpose

This thesis researches the impact of investor sentiment and attention on price formation of Bitcoin under the framework of behavioural finance. The primary focus is to investigate whether the Asian investors' sentiment affect Bitcoin pricing differently from the European investors' sentiment and vice versa.

Besides investor sentiment and attention, the relation between Bitcoin and the traditional asset classes has also interested many researchers (MacDonell, 2014) (Dyhrberg, 2015) (Dyhrberg, 2016) (Bouri, et al., 2017). Bitcoin has disrupted the existing monetary market and posed great challenges to regulators. The interaction between this new system and the old system may affect the pricing of both systems. I thus add traditional assets including VIX, FTSE 100, Shanghai A-share and gold to my regression models.

1.3   Contribution

Different from many recent Bitcoin related researches which focus on investor sentiment aggregated from social media platforms (Zhang, et al., 2011), (Karabulut, 2012), (Abraham, et al., 2018), (Nasekin & Chen, 2019), (Valencia, et al., 2019), I collect news articles published by two news agencies, namely SCMP and FT. I have not found researches using data from these two sources to investigate Bitcoin pricing. News articles are inherently different from short discussions posted on social media platforms as already discussed in section 1.1.1. Researching news data may uncover fresh findings.

The other novelty of this thesis is to study the continental discrepancies between the sentiment of Asian investors and that of European Bitcoin investors. For the robustness of the results, I also employ an alternative approach -- the search-based attention measure, Google Search Volume Index (SVI), to cross validate my hypothesis.

Further Granger-causality tests are conducted between search-based and news-based attention measures to understand the lead-lag effects of these two attention measures and their interaction with Bitcoin and news sentiment.

1.4   Delimitation

Due to the limitation of the SCMP web portal, I am unable to acquire a complete set of SCMP news of the sampling period. Hence, the inferences drawn from the SCMP dataset is treated with caution.

The relation between public sentiment and news sentiment is complex and debatable. It is unclear to what extent textual sentiment causally relates to investor sentiment. As Kearney et al. pointed out that the "fundamental difference between investor sentiment and textual sentiment is that the former captures the subjective judgments and behavioural characteristics of investors, while the latter can include the former but also includes the more objective reflection of conditions" (Kearney & Liu, 2014). Empiricists, such as Tetlock observed that "high values of media pessimism induce downward pressure on market prices" and noticed the linkage between media content and the behaviour of individual investors (Tetlock, 2007). Interaction between mass media and public opinions is intricate and complicated. This specific area is not only intriguing but also of paramount significance, because the hypotheses that news sentiment affects Bitcoin pricing is built upon the assumption that news sentiment is an accurate proxy of investor sentiment. However, limited by the scope of this thesis, I do not attempt to challenge the validity of this assumption but just directly use it to build my hypotheses.

The sentiment analysis is obtained using Loughran's textual analysis script[1]. This script was originally designed to analyse the polarity of 10-K documents. Thus, the performance on news-based texts is not thoroughly studied. Furthermore, the count-based sentiment analysis algorithm only uses simple terms to express sentiment. Complex elements, such as cultural factors, linguistic nuances and differing contexts are too difficult to quantify into a simple sentiment measure. Even humans often disagree with each other on the sentiment of text, not to mention the precision of computer algorithms. But before more sophisticated algorithms are available, my research can only assume the existing approach is good enough to measure sentiment. However, future researches shall follow the development of textual analysis and always adopt to a better approach if available.

---

[1] Downloaded from https://sraf.nd.edu/textual-analysis/code/

Whether SCMP and FT capture the sentiment of Bitcoin investors is not entirely clear. Even though these two news sources are influential as news providers, Bitcoin investors may not be the readers of news. This mismatch likely brings in noise to the final results.

The relation between investor sentiment and investor attention is only superficially explored in this thesis. Attention is proxied through news volume, a much simpler and direct measure. Intuitively, when news volume increases, sentiment is expected to attenuate. But their low correlation concludes otherwise. Whether this is due to the inaccuracy of sentiment measurement or some other reasons is unclear.

The endogenous factors, such as supply and demand, transaction data on the blockchain, hash rate, are not studied. These factors are also very crucial to Bitcoin pricing. Future studies shall consider including these factors.

1.5   Structure of this thesis

This thesis is arranged as follows. Section 2 explains the theoretical framework this study is based upon. Section 3 reviews the previous relevant literatures about sentiment analysis, investor attention and their impact on traditional assets and also on Bitcoin. Section 4 presents the data collection process and data sample construction and provides an overview of the data sample. Section 5 elaborates the modelling process, shows the regression results. Section 6 conducts Granger causality analysis. Section 7 concludes this study and discusses future work.

2   THEORETICAL FRAMEWORK

"Beauty is in the eyes of beholders", so is value.

The divergence between the fundamental value and the market price has been a central issue in asset pricing, especially in the emergence of a whole new asset class, such as cryptocurrencies. Can people reach a consensus about the fundamental value of Bitcoin?

Even though the designer(s) of Bitcoin intended to create a new replacement monetary system free of inflation, transacted through a payment system in elimination of intermediaries, Bitcoin's high volatility and low usability unfortunately restrain its currency function. Cheah et al. even conclude that the long-term fundamental value of Bitcoin is zero (Cheah & Friy, 2015). The supply and demand equilibrium model become unsuitable when applied to Bitcoin pricing. It is even still debatable if Bitcoin is investment, commodities, or currencies (Grinberg, 2012). Without a commonly accepted valuation method, Bitcoin pricing is like the wild west.

In this chapter I review the major existing asset pricing theories in an attempt to establish a theoretical framework for my research about Bitcoin price formation. Furthermore, I also dive into the technical foundations of sentiment analysis to enhance the robustness of my research.

2.1   Neoclassical finance versus Behavioural finance

Neoclassical finance is built around clean and unified financial theories including the portfolio principles (Markowitz, 1952), the arbitrage principles (Modigliani & Miller, 1958), the capital asset pricing theory (Sharpe, 1964) and the option-pricing theory (Black & Scholes, 1973). The Efficient Markets Hypothesis assumes market participants are rational (known as "arbitrageurs") and a security's price equals its fundamental value. Deviations from fundamental values are quickly arbitraged away by "arbitrageurs". As summarized by Ramiah et al. in their literature review about asset pricing theories, neoclassical finance believes that "financial markets react quickly to new information; prices follow a random walk process resulting from the random arrival of information; and no investor can consistently earn abnormal return in excess of what is consistent with risk" (Ramiah, et al., 2015). The neoclassical finance replies more on statistical approach based on Von Neumann Morgenstern theory and Bayesian techniques.

Unfortunately, the reality is much messier. Neoclassical finance fails to explain the persistent market anomalies. Market participants are not uniformly rational. Instead of solely relying on mean-variance configuration, noise traders may make decisions based some heuristic factors such as taste, preference, sentiment and other psychological factors. Different from neoclassical finance, behavioural finance accounts for market inefficiency caused by human errors and also acknowledge the impact of noise traders as active market participants whose participation gives rise to market anomalies. Behavioral finance applies social psychology to explain market anomalies. As pointed out by Ramiah et al., social psychology focuses on interpersonal behaviour and the role of social forces in governing behaviour (Ramiah, et al., 2015).

2.2   Prospect theory

As early as 1912, Selden already observed the tight connection between stock price and the mental status of market participants (Selden, 1912). But the most renowned contribution in this field belongs to the prospect theory developed by Tversky and Kahneman. In contradiction to the expected utility theory, in prospect theory, people prefer outcomes that are obtained with certainty instead of outcomes that are merely probable, irrespective of the expected utility (Kahneman & Tversky, 1979). In other words, people react asymmetrically between potential losses and potential gains, which often leads to overreaction to small probability events and underreaction to large probability events.

In another paper of theirs, Tversky and Kahneman studied how people form such "irrational" preferences. Choices are presented to test subjects in a way that highlights the positive or negative aspects of the same decision. If people were rational, different representations of the same decision should yield the same preference. However, Tversky and Kahneman observed that variations in the framing of decision problems produce systematic violations of invariance (Tversky & Kahneman, 1986). Instead of considering the maximum utility, people make decisions based on the potential gain or losses relative to their reference points, e.g., their initial wealth.

The framing effect suggests that the reference point of an individual is subjected to manipulation. Different representations of an essentially identical problem can change people's choices. As observed by Scheufele and Tewksbury, a framing effect occurs when audiences pay substantial attention to news messages (Scheufele & Tewksbury, 2007). Furthermore, Chong and Druckman's study finds that repetition enforces the framing

effect on less knowledgeable individuals, "whereas more knowledgeable individuals are more likely to engage in systematic information processing by comparing the relative strength of alternative frames in competitive situations". However, framing effects are formed less likely on established issues and among knowledgeable people who are aware of the central considerations on the issue (Chong & Druckman, 2007).

2.3 Noise trading

A number of researchers refer to market sentiment effects as noise trading. Noise trading has been identified as a major source of volatility, known as "noise trader risk". Extensive experiments conducted by cognitive psychologists reveal that systematic biases arise when people form beliefs (Barberis & Thaler, 2003). Many researchers observed that "securities that have had a long record of good (bad) news tend to become overpriced (under-priced) and have low (high) average returns afterwards" (Barberis, et al., 1998).

De Long et al. proposed a model to quantify noise trader risk (De Long, et al., 1990). Under their model, the demands for the risky asset are proportional to its perceived excess return and inversely proportional to its perceived variance. When noise traders overestimate expected returns, they demand more of the risky asset than sophisticated investors do; when they underestimate the expected return, they demand less.

Shefrin and Statman developed a behavioural asset pricing model (BAPM) and added the noise trader risk to the CAPM beta (Shefrin & Statman, 1994). According to their model, noise traders also affect the volume of trade due to price efficiency.

2.4 Measurement of investor sentiment and attention

Academia have devised at least four main approaches to measuring investor sentiment and attention: financial market-based measures, survey-based sentiment indices, textual sentiment and internet search data, broadly categorizing into direct measures and indirect measures.

Financial market-based measures are usually indirect measures proxied through including trading volume, dividend premium, closed-end fund discount, the number and first day returns on IPOs, and VIX (Baker & Wurgler, 2007). But the problem with these indirect measures to quote Da et al.: "they have the disadvantage of being the equilibrium outcome of many economic forces other than investor sentiment" (Da, et al., 2014). This is also a general issue of using indirect measures, the deviation and distortion from the origin.

The survey-based sentiment index is a direct measure since the information is usually collected through interviews and questionnaires to consumers or investors. Solt and Statman used the Bearish Sentiment Index (the ratio of the number of bearish advisers to the number of all advisers expressing an opinion) and the survey data published by Investors Intelligence (Solt & Statman, 1988). Verma and Verma use both the survey data of American Association of Individual Investor and Investors Intelligence (Verma & Verma, 2007). But most survey data come at a lower frequency, weekly or monthly, rarely daily. The truthfulness and accuracy also depend largely on respondents' motives and incentives provided (Singer & Ye, 2013).

2.4.1 Textual sentiment analysis

Textual sentiment analysis as a subdivision of Natural Language Processing can be roughly divided into lexicon-based methods and machine-learning methods. As summarized in Maynard and Funk's paper, lexicon-based methods rely on a sentiment lexicon, a collection of known and pre-compiled sentiment terms; machine learning approaches make use of syntactic and linguistic features (Maynard & Funk, 2011).

The machine learning approaches have gained popularities in recent decade. Sentiment extractions built using supervised methods can usually achieve high accuracy in its targeted domain (Boiy & Moens, 2009). But this approach requires large amount of labelled data, which usually need engagement of human experts. Acquiring such data is a time consuming and expensive process. Moreover, analysing longer texts such as news articles requires extensive computing power. Another constraint of machine learning approach is the limited applicability of the models. One trained model is usually only applicable to one specific task. According to Aue and Gamon, the good performance of supervised methods drops significantly if applied to a different domain (Aue & Gamon, 2005). This thesis does not use machine learning approach because I do not have the resources to acquire labelled data.

The lexicon-based approach simply calculates the proportion of semantic orientation of words in a document based on some precompiled word lists. Compiling such lists also rely on human experts. But different from the machine learning approach, once such precompiled word lists are constructed, they can be applied to a wide variety of texts. Furthermore, the human engagement of annotating texts needed for machine learning models is much more than that needed for compiling tonal word dictionary.

The Harvard's General Inquirer (GI) is a commonly used external word list derived from applications in psychology and sociology. GI contains 182 categories of sentiment including positive, negative, strong, weak, active, passive etcetera. Finance and accounting researchers generally focus on the Harvard Psychosociological Dictionary, specifically, the Harvard-IV-4 TagNeg file which is a group of negative words. However, Loughran et al. show that the TagNeg list performs poorly for business applications, e.g., "mine", "liability" and "vice" are negative words that occur frequently in business without any negative implications (Loughran & McDonald, 2011). Therefore, they identify a new set of tonal words to measure sentiment in the context of financial applications.

2.4.2 Internet search behaviour

Simon concludes that people start their decision-making process by gathering relevant information (Simon, 1955). In the digital era, information gathering often involves searching online sources. Since 2004, Google has begun to provide aggregated information on the volume of queries for different search terms, via the publicly available service Google Trends.

Search volumes data provides valuable information in predicting investor attention and market returns. Mondria et al. found empirical evidence on the interaction between international asset holdings and investor attention measured through American Online "search/click-through" records (Mondria, et al., 2010). Da, et al. propose Google SVI as a direct measure of investor attention (Da, et al., 2011). Choi and Varian have shown that Google Trends data can be linked to automobile sales, unemployment claims, travel destination planning and consumer confidence (Choi & Varian, 2012). Preis et al. found patterns out of the extensive behavioural search data that notable drops in the financial market are preceded by periods of investor concern (Preis, et al., 2013).

The predictive power of search data is also recognized in Bitcoin. Kristoufek finds a striking positive correlation between Bitcoin price and the searched terms measured by Google Trends and Wikipedia (Kristoufek, 2013). Abraham et al. construct Bitcoin daily price prediction model using SVI and tweet volumes with high accuracy (Abraham, et al., 2018). Figà-Talamanca and Patacca analyze find that market attention measured by SVI has an impact on Bitcoin returns (Figà-Talamanca & Patacca, 2018). Bleher and Dimpfl, however, find that returns are not predictable while volatility is predictable to some extent using SVI (Bleher & Dimpfl, 2018).

3 REVIEW OF PREVIOUS LITERATURES

Literatures are reviewed from two broad categories:

- Choice of data used to proxy investor sentiment and attention, primarily from three sources including social media platforms, search behaviour data, news-based sentiment data.

- Textual analysis techniques including machine learning approach and lexicon-based approach.

Some of the earlier researches in the field of the sentiment-based Bitcoin pricing hypothesis utilize Wikipedia search data and Google Trends to proxy investor sentiment (Kristoufek, 2013) (Florian, et al., 2014) (Garcia, et al., 2014). Search data is a very good proxy for investor attention. Researches using search data do not need to concern the textual mining techniques.

As time evolves, more abundant data relevant to Bitcoin becomes available. Due to the large user base and convenient APIs for efficient data queries, many of the researches resort to social media platforms, including Twitter (Kaminski & Gloor, 2014) (Lamon, et al., 2017) (Abraham, et al., 2018), StockTwits (Chousa, et al., 2016) (Nasekin & Chen, 2019) or even across both Twitter and StockTwits (Dong, et al., 2019) to aggregate investors' sentiment towards Bitcoin. Data from social media platforms is abundant and timely, which is ideal for tracking and even predicting the short-term movement of Bitcoin price.

Another hunting ground for investor sentiment is news articles. There are relatively fewer researches using news-based sentiment, which is likely due to the difficulties of data queries, the burden of computation and the low frequency of news-events relative to microblog messages, especially when the primary interest lies in the short-term prediction of Bitcoin prices.

Depending on the choice of data sources, some researches use lexicon-based textual analysis (Kaminski & Gloor, 2014) (Abraham, et al., 2018), for example simply counting the tonal words or using open source packages specifically developed for analysing sentiment of social media contents. Others employ machine learning approaches (Lamon, et al., 2017) (Nasekin & Chen, 2019). But the machine learning approach is

usually limited to analysing short texts from social media platforms or news head-lines due to the computation constraints.

In section 3.1 and 3.2, I present two influential researches which provide technical foundation for the sentiment analysis and theoretical framework of attention induced price pressure hypothesis. In the last three sections, I present in more details of three representative papers in the area of the sentiment-based Bitcoin pricing hypothesis.

3.1 When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks (Loughran & Bill 2011)

This is one of the most important literatures which inspires and facilitates my research. Loughran and McDonald constructed a new dictionary (consists of six lists: Fin-Pos, Fin-Neg, Fin-Unc, Fin-Lit, MW-Strong and MW-Weak) based on the commonly used Harvard Psychosociological Dictionary (HPD) for sentiment analysis, because they noticed the HPD misclassified words when gauging tone in financial documents. To validate their findings, they further applied the pre-compiled word list to data samples selected out of all 10-Ks and 10-K 405 files over 1994 to 2008 and examined the impact of the sentiment of these financial documents on stock returns.

They use corporate disclosures to capture sentiment from management, the insiders who know most about their firms. The raw files are then parsed into vectors of words and word counts which are used to calculate the percentage of tonal words over total number of words in a document. The sentiment of words is categorized into six lists: negative, positive, uncertainty, litigious, strong modal words and weak modal words.

Controlling for firm size, book-to-market, share turnover, pre-file date Fama-French alpha and a NASDAQ dummy, the multivariate analysis suggests a higher proportion of negative words (classified by the Fin-Neg list) are associated with lower excess returns with statistical significance. The Harvard list (H4N-Inf) on the other hand is not significantly related to the file date excess returns. Additionally, besides the simple percentage measure of tonal words, they also devised a term weighting scheme which considered the importance of a term within a document and the entire corpus. This more sophisticated approach can mitigate the noise of word misclassification. Loughran and McDonald share their Python script for text analysis on the internet which has been used by my thesis. Their achievement approves the importance of domain specific adaptation when replying on outside results.

3.2 In search of attention (Da, et al., 2011)

In 1987, Merton proposed the theoretical framework in which limited investor attention can affect asset pricing. But it has been challenging for empiricists to identify a direct measure of investor attention. Da, Engelberg and Gao propose the aggregate search frequency (SVI) in Google as a direct measure of investor attention because of the dominant share of Google in the search engine market and the revealing searching behavior of individuals.

The research finds that SVI is only weakly correlated with alternative attention measures including extreme returns, turnover and news-based measures, indicating the discrepancy between SVI and existing proxies as attention measures. Further lead-lag relation analysis using VAR reveals that SVI leads the other three attention proxies with statistical significance. This suggests that SVI can be a timelier measure of attention than extreme returns or news.

Moreover, cross-sectional regression against different order sizes suggests that SVI especially captures retail investors attention, which results in buying pressure that pushes stock prices up temporarily but almost completely reversed in a year. The effect of retail attention on asset prices is also observed through the IPO returns. SVI has strong incremental predictive power for first-day IPO return and long-run underperformance among IPO stocks with high first-day returns.

This paper provides empirical evidence for the attention-induced price pressure hypothesis. Considering the speculative nature of Bitcoin pricing, the price pressure hypothesis may provide a suitable theoretical framework for my research. Unfortunately, there has not been an established way of identifying retail investors in Bitcoin transactions. Tracking the transaction data on the Bitcoin blockchain may be one way to categorize the types of Bitcoin investors, which has to be left to my future research to explore.

3.3 Using sentiment analysis to predict interday Bitcoin price movements (Karalevicius, et al., 2018)

The data collection methodology and textual analysis approach of this paper are very similar with mine, which is also through web scrapping and lexicon-based techniques. However, the purpose of their research is very different, which justifies their choice of the news sources (exclusively Bitcoin oriented), including Bitcoin expert media, CoinDesk, Cointelegraph and NewsBTC. News stories on websites dedicated to Bitcoin

cover most publicly available Bitcoin-related information and hence, is a good proxy for a media sentiment surrounding the Bitcoin.

Karalevicius et al. pre-processed the raw texts into tokenized and tagged words using the Natural Language Toolkit (a suite of libraries and programs for symbolic and statistical natural language processing for English), and further obtained the correct lemma through WordNet Lemmatizer. Next they conducted the sentiment analysis using both Harvard Psychosocial Dictionary and Loughran and McDonald's word list. To evaluate the performance of the two dictionaries, they proposed a trading strategy as below:

- An investor assesses the news every day, and at the end of the day he makes the decision to act or not act (no shorting allowed).

- In case of positive signals, he will go long the first day, whereas negative signals will induce a short position.

- Do the exact opposite for the next two days to catch the corrective price movement.

- Revert the position again hoping that increased market liquidity steers the price to a more long-term trend.

Next they simulated the market conditions with different durations and starting points. The interday price pattern suggests that after the publication of an expert news story, the price first goes in the direction of the sentiment, but the market overreacts a little. As a result, the price makes a corrective movement. A trader who fully exploits all price movements cannot achieve abnormal returns because of the transaction costs and the elevated risk of the strategy. It is, however, inconclusive whether Loughran and McDonald's dictionary outperforms the Harvard list according to their results.

This study focuses more on the language processing techniques. The simulation of the trading strategy is introduced primarily to evaluate the performance of the sentiment analysis approaches. This is a black-box approach solely to capture the phenomenal characteristics between news-based sentiment and Bitcoin price movement, disregarding the causal effect between media coverage and Bitcoin pricing. But the data processing and textual analysis strongly resemble my methodology. Their results provide empirical evidence for the notable impact of news sentiment on Bitcoin pricing in the short-term. This study enhances the validity of the technical foundations of my

methodology. The more sophisticated textual processing shall also be considered in my future studies.

3.4 Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis (Abraham, Higdon, John, Juan 2018)

Abraham, Higdon, John and Juan research the relation between cryptocurrency related tweets and Bitcoin prices. Their findings, however, suggest the sentiment of tweets cannot serve as a reliable proxy for future cryptocurrency demand, instead that the volume of tweets and Google Trends data are more robust predictors of bitcoin price direction.

First, they collect posts from Twitter through APIs and then analyse the sentiment of these posts using VADER (Valence Aware Dictionary and sEntiment Reasoner). VADER is an open source textual analytic tool specifically designed for social media contents. It does not only classify text as positive, negative, or neutral but also measures the intensity of words and generates a compounded sentiment score for each piece of text.

Next, they obtain the Google trends data which comes as a search volume index (SVI) calculated by dividing each data point by the total searches within a geographic region and time range. Google also provides APIs for requesting daily trends data.

Finally, they investigate the correlations between tweet sentiment and price; tweet volume and price; SVI and price. The VADER sentiment scores suggest half of the tweets are neutral. But even after filtering out the objective tweets, the emotional tweets do not present a consistent pattern against price changes in a 90-day period. However, the correlation analysis between SVI and price changes and between tweet volumes and price changes result in Pearson Rs of 0.817 and 0.841 and p-value of 0 and 0 respectively. Furthermore, their multiple linear regression using SVI and tweet volumes as input variables predict Bitcoin closing daily price with high accuracy.

This paper may have fallen short for not doing cross validation, but the strong correlation between tweets volume and Bitcoin price is compelling by itself, which is consistent with my observation in this thesis between news volumes and prices.

3.5 Deep Learning-Based Cryptocurrency Sentiment Construction (Nasekin & Chen, 2019)

Different from Loughran and McDonald's approach which uses bag-of-words and term weighting scheme, Nasekin and Chen use machine learning techniques, more specifically

long short-term memory (a variant of recurrent neural network), to analyse sentiment. The notable benefit of using machine learning on textual analysis is that the context around a word can be "remembered", which improves the accuracy of sentiment evaluation. However, we must not neglect the two major drawbacks of supervised learning: it needs large amount of labelled training data which is time consuming and expensive to acquire; the cost of computation is high for training high-dimensional classifiers.

The aforementioned hurdles are mitigated by using data from StockTwits a microblogging platform for traders and investors. Users of StockTwits can express their sentiment by labelling messages as "Bearish" (negative) or "Bullish" (positive). Such labelled data are ideal training candidates for supervised learning. Moreover, conversations are organized around "cashtags" which help identifying cryptocurrency related contents. The same as Twitter, text messages on StockTwits are short comparing to news articles which require much more computing capacity. Hence, the choice of StockTwits mitigates the two major hurdles of using supervised learning.

Four sentiment indices are then constructed using two different approaches: a "seed" lexicon of terms and the augmented context-specific terms; a trained RNN model to predict unlabelled messages. The four indices are regressed respectively against the logarithmic returns of the cryptocurrency index "CRIX" (Trimborn & Härdle, 2018). The best significance results are achieved by the sentiment index using lexicon expansion. Therefore, including domain-specific information appears to improve the predictability of returns. Additionally, to test the generality of the indices, they also inspect the impact on more traditional assets, Apple stock. But the predictability diminishes. Finally, they investigate the impact on the volatility of CRIX using the integrated GARCH method.

The research methodology and the results of this paper is inspired and of meaningful comparison to my research. Despite using different approach for sentiment analysis from mine, this paper presents a good framework of researching the relation between sentiment and cryptocurrency prices. The comparison groups (regressions for the four sentiment indices) not only enrich the content of the research but also reveal the importance of domain-specific information which refers to the financial aspects such as market activities and transactions instead of the technical aspects. This is not surprising considering how uncertain the technical aspects still are at the current phase. Therefore, establishing a reliable mechanism to construct sentiment index helps explain the price movement.

4 DATA

4.1 Choice of data sources

News articles published between 22nd April 2013 and 30th June 2019 are collected from the web portals of South China Morning Post (SCMP) and Financial Times (FT). The starting sampling period is determined based on the appearance of the first Bitcoin news on these two sources.

SCMP is a Hong Kong English-language newspaper. It is chosen because tonal word lists for Chinese language is not available and there are no other electronically available Chinese English-news providers. The cryptocurrency related activities including mining, trading and ICOs, have been active in China even after the 2017 regulation. Chinese investors' sentiment likely has considerable impact on Bitcoin pricing.

FT, one of the world's oldest and most respected news publications, plays a key role in shaping public opinion and investor reaction. Dyck et al. find that the Financial Times has more credibility and influence than other news sources. One more article in the Financial Times increases the probability of reversing a corporate governance violation by five percentage points (Dyck, et al., 2008). Hence, FT is chosen to proxy the sentiment of European investors.

	SCMP	FT
china/asia	666	308
uk/europe	144	680

As shown in the left, among the Bitcoin news, 666 items from SCMP have keyword 'china' or 'asia' and only 308 news from FT mentioned 'china' or 'asia'. The result is reversed for keyword 'uk' and 'europe' for SCMP and FT respectively. Evidently SCMP covers more news events from China and Asia, while FT covers more activities happening in UK and Europe. The reginal focus of these two sources is obvious. The choice of using SCMP to proxy Asian investor's sentiment and FT to proxy European investors' sentiment should be feasible.

4.2 Data collection and cleaning

FT provides APIs for headline license requests. Users can specify news published dates in the HTTP request, which guarantees the precision and completeness of search results. URIs to news articles are then obtained through headline requests on a yearly basis.

SCMP does not provide APIs for developers. I have to obtain URIs through both the SCMP's RSS feeds and keyword searching. This is because RSS feeds do not contain a complete set of published news. To have a better coverage, I also gather the query results using different keywords. But the search interface through SCMP's web portal is very limited, without any filter functionalities, such as time constraints; one can only search with certain keyword. Different from FT, SCMP covers more general topics about many non-financial aspects of the world. There are eight subdivisions under news tag, among which includes business, economy and tech. And in each of these divisions, there are different sections. I use the names of these sections and also words often used in business context to gather relevant news.

Table 1 contains the keywords used for searching news in three categories. Because the first Bitcoin news occurred in 2013, so news older than 2013 are filtered out. Unfortunately, such heuristic approach cannot guarantee the completeness of the news sample. Without SCMP provided APIs or access to SCMP's database, one cannot ensure the completeness of the search results due to the limited searching interface on SCMP's web portal. Baker, Bloom and Davis also used keywords to query news from SCMP (Baker, et al., 2016). The difference is that they have access to the database ProQuest which provides much more versatile and precise searching interface.

Table 1 **Keywords used to search for news in SCMP**

Category	Keyword
Business	price, market, invest, buy, sell, business, company, bank, finance, property, start-ups, gear, e-commerce, capital, stock, bond, currency, commodity,
Economy	economy, policy, regulation
Tech	tech, crypto, blockchain, apps, bitcoin

Once I obtain all the URIs from both web sites, the contents of each article are retrieved through separate HTTP requests. The replies of these HTTP request are raw HTML pages which have to be cleaned by removing all HTML tags and Javascripts. With the help of Python scripts, the contents of each news article are written into a plaintext file together

with the date of publish and the title. The various scripts used during web scraping are developed in shell commands, Perl and Python.

4.3 Overview of the news data samples

There are 269569 news articles retrieved from FT between 22rd April 2013 and 30th June 2019. As shown in Table 2 the average level of the negativity (percentage of negative words) and the positivity (percentage of positive words) of all news are fairly even throughout the years, around 2.5 and 1 respectively. The average negativity of Bitcoin news published by FT decreased from 2.44 in 2017 to 2.26 in the first half of 2019. But the volume of Bitcoin related news is almost tripled from 147 in 2016 to 391 in 2017 and continued the trend to 622 in 2018.

As explained above, I only managed to acquire a partial set of SCMP news. Especially the earlier years from 2013 to 2016, there are only hundreds of news. SCMP's web service does not return all entries when the search result exceeds certain limit. The incompleteness of the data needs to be borne in mind when comparing against FT sample. But the query results for Bitcoin news (news contents containing at least one occurrence of "bitcoin" keyword) is reliable. Notably the Bitcoin news volume increased drastically in 2017 and 2018 which were also the years Bitcoin prices surged to historical high. And meanwhile the average negativity of Bitcoin news decreased from 2.57 in 2016 to 2.43 in 2017, to 2.38 in 2018 and continued to decrease to 2.24 in the first half of 2019.

According to Table 2, Bitcoin news from SCMP appear to be in average more negative than these from FT, especially during 2014 and 2015 with the percentage of negative words over 3.1 and 3.64 respectively. But FT has published more news annually and 588 more news in total about Bitcoin since 2013 than SCMP has, which is not surprising since FT is a larger news provider and more specialized in financial news.

Another interesting observation in Table 2 is that the mean negativity of news is consistently larger than the mean positivity both for Bitcoin news and all news. Figure 1 is the histogram of Bitcoin news sentiment (the percentage of positive words subtracts the percentage of negative words). Apparently, there are more news with negative sentiment scores than those with positive scores. In other words, Bitcoin related news in average use more negative words than positive words. This observation is also present in Loughran et al.'s research (Loughran & McDonald, 2011), where in average the sentiment of 10-K documents is negative. Unfortunately, Loughran et al. do not offer an explanation about this result.

Table 2 **Descriptive statistics of news articles from FT and SCMP**

		2013	2014	2015	2016	2017	2018	2019
FT	Total news	31687	44866	43533	40968	39057	48747	20711
	Bitcoin news	71	145	155	147	391	622	147
	Mean negativity	2.51	2.47	2.47	2.53	2.50	2.63	2.72
	BTC mean negativity	2.66	2.69	2.63	2.4	2.44	2.36	2.26
	Mean positivity	1.10	1.08	1.07	1.06	1.06	1.04	1.03
	BTC mean positivity	0.94	0.97	0.98	1.07	1.03	0.99	1.18
SCMP	Total news	120	202	155	892	9259	29040	24275
	Bitcoin news	59	119	61	87	293	384	87
	Mean negativity	2.36	2.77	2.83	2.49	2.13	2.44	2.68
	BTC mean negativity	2.69	3.1	3.64	2.57	2.43	2.38	2.24
	Mean positivity	0.95	0.92	0.87	1.06	1.16	1.01	1
	BTC mean positivity	0.74	0.79	0.7	1.07	1	0.88	0.91

This phenomenon brings up two questions: whether this is because news sentiment is really in average more negative than positive towards Bitcoin; or this result is just due to the sentiment analysis algorithm but not inherent to any particular topics. To better understand the result, I also plot the histogram of the sentiment measures for all news collected from SCMP and FT respectively. As shown in Figure 2, irrespective of topics, the sentiment of all news from these two sources is also negative in average. Hence, it is not unique of Bitcoin news being more negative than positive in average. Since I already assume that Loughran et al.'s tonal word dictionary is reliable, the conclusion is likely because that in average news is negative. And comparing Figure 1 against Figure 2, the Bitcoin news are not more negative than the sentiment of average news.

Figure 1   Sentiment of Bitcoin news from FT and SCMP

Histogram of Bitcoin news sentiments published by FT and SCMP. Red is for news published by FT. Blue is for news published by SCMP. The X-axis is the difference of the percentage of positive words subtracting that of negative words. Y-axis is the number of Bitcoin news. The sampling period is from 22nd April 2013 to 30th June 2019.
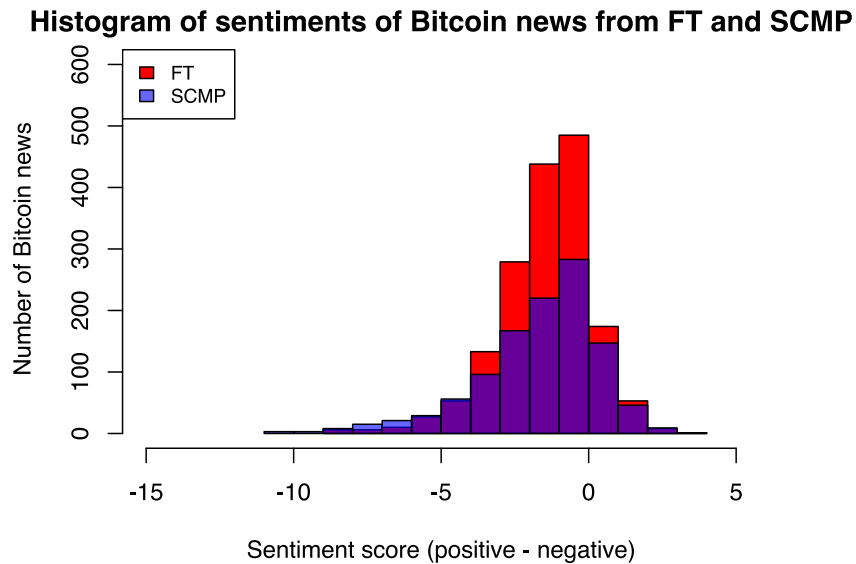
Figure 2   Sentiment of all news from FT and SCMP

Histogram of news sentiments of all news samples collected from FT and SCMP from 22nd April 2013 to 30th June 2019. Red is for news published by FT. Blue is for news published by SCMP. The X-axis is the difference of the percentage of positive words subtracting that of negative words. Y-axis is the number of Bitcoin news. The sampling period is
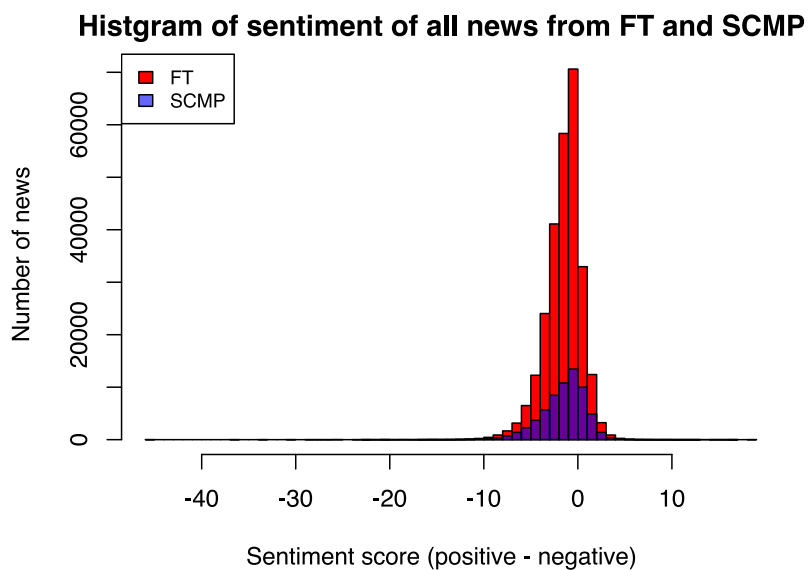
Figure 3 plots the monthly average Bitcoin prices and the monthly news volume from FT and SCMP for 22nd April 2013 to 30th June 2019. The red line is the monthly average Bitcoin price which spiked from the end of 2017 through 2018. The black and blue lines are the monthly news volume from FT and SCMP respectively, which tightly follow the fluctuation of Bitcoin price. There seems to exist strong correlation between the Bitcoin news volume and Bitcoin prices.

Figure 3   Monthly Bitcoin prices and Bitcoin news volume from FT and SCMP

Monthly average Bitcoin price and news volume from FT and SCMP
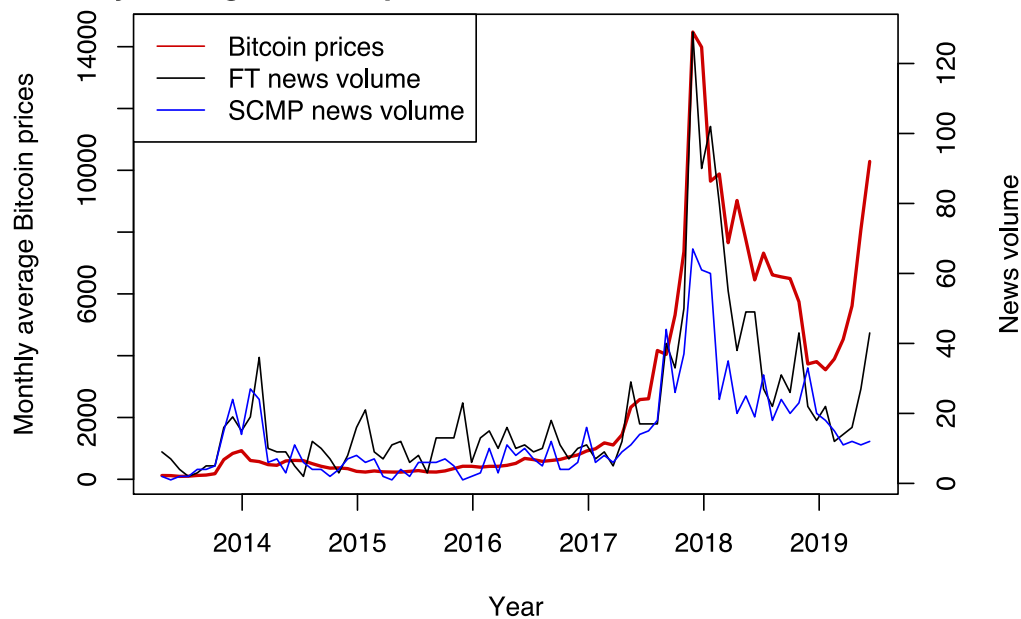
The thicker red line is the monthly average Bitcoin price. Black line is the number of news published by FT. Blue line is the number of news published by SCMP. The X-axis is the time horizon. The left Y-axis is the monthly average Bitcoin price. The right Y-axis is the monthly number of Bitcoin news. The sampling period is from 22nd April 2013 to 30th June 2019.

Figure 4 plots the monthly average Bitcoin prices and the monthly news sentiment (the percentage of positive words subtract the percentage of negative words) from FT and SCMP for 22nd April 2013 to 30th June 2019. The black and blue lines are the monthly news sentiment from FT and SCMP respectively. The correlation between sentiment and Bitcoin price is not as pronounced as that between news volume and price. It appears that sentiment in the early years is more extreme and negative, but gradually soothes out in later years.

Hypothetically, negative news should caution investors from rushing into Bitcoin and lead to price correction. News like the crackdown of Silk Road and later the default of Mt. Gox indeed caused a significant price drop. But the subsequent price recovery

renders this hypothesis invalid. No matter the default of Mt. Gox or association with the notorious dark web Silk Road seem to be just a singularity that affected a subset of users and are apparently not serious enough to lead to significant price corrections.

Appendix 7 and Appendix 8 list the news articles having the highest percentage of negative words from the FT and the SCMP news samples respectively. The most negative news articles are all about dubious crimes, theft or fraud around Bitcoin, but not about the underlying technologies of Bitcoin. The unfortunate association with criminal activities may actually on the other hand prove the desirability of Bitcoin. Florian et al. observe that serious negative events, do not lead to significant price corrections and conclude that Bitcoin users seem to be positively biased towards Bitcoin (Florian, et al., 2014). Even though the contents of these news are negative, the reception from Bitcoin investors may actually be positive and enhancing.

Figure 4   Monthly average Bitcoin prices and news sentiment from FT and SCMP

The thicker red line is the monthly average Bitcoin price. Black line is the sentiment of news published by FT. Blue line is for SCMP samples. The X-axis is the time horizon. The left Y-axis is the monthly average Bitcoin price. The right Y-axis is the sentiment of Bitcoin news (percent of positive words – percent of negative words at monthly frequency). The sampling period is from 22nd April 2013 to 30th June
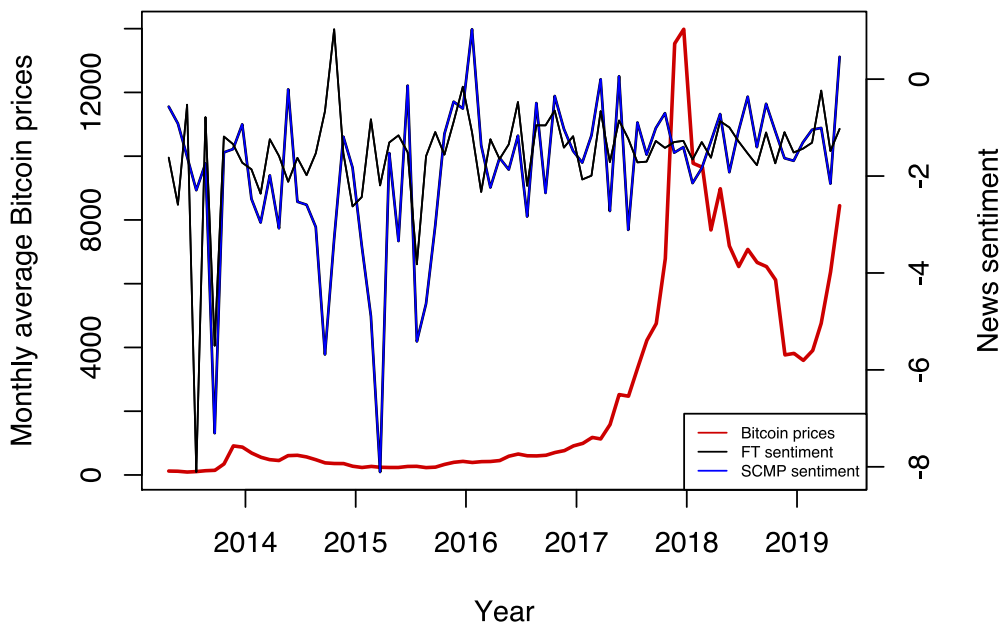
Appendix 9 and Appendix 10 list the news articles having the highest percentage of positive words from the FT and the SCMP news samples respectively. Different from the most negative news list which have some overlapping contents from both sources, the

contents of the most positive news are fairly different between SCMP and FT. Appendix 11 list the thirty most frequent negative and positive words used by the Bitcoin news from both SCMP and FT. Among which 'against', 'laundering', 'crisis' is the three most frequently used negative words, and 'good', 'best', 'better' is the top three positive words.

5 METHODOLOGY AND RESULTS

The value of Bitcoin is hypothetically driven by two factors: the level of public interest over time, measured by news sentiment and new volume; the level of investor confidence in traditional markets proxied through gold, VIX, FTSE 100 indices and Shanghai A-share indices. Data samples used for OLS regression analysis are constructed at weekly and monthly frequencies. Ideally it would provide more timely information if data samples were taken at a daily basis. But news about Bitcoin from the chosen two news sources is not available at a daily frequency. In the following sections, I introduce the variables used in the regression analysis and elaborate the reasoning behind my choices. Detailed variable definitions can be found in Appendix 1.

5.1 Dependent variables

My primary tests examine Bitcoin return and Bitcoin price relative to news sentiment respectively. The natural logarithms of weekly and monthly Bitcoin return are the dependent variables, as well as the weekly average Bitcoin price. The dramatic fluctuation of Bitcoin prices has made eye-catching sensational headlines on media for many times. Investors' sentiment is influenced by the ups and downs of Bitcoin prices. Moreover, as already discussed in section 2 about the challenges on pricing Bitcoin, regressions using prices may provide more insights about Bitcoin price formation.

5.2 Independent variables

5.2.1 *News-based measures*

The sentiment of a news is measured by the percentage of positive and negative words. Daily positivity and negativity are aggregated by averaging respectively the percentages of positive and negative words of all the news published on the same day. Weekly sentiment measures are then constructed based on the daily data. The sentiment score and news volume are set to zero for weeks without any Bitcoin news. Sentiment score and news volume are the independent variables whose impact on Bitcoin is of primary focus of this thesis.

The dummy variable of news origin is added to identify the publisher of a news for cross-sectional regressions. The sentiment score is defined as the percentage of positive words minus the percentage of negative words, and then subtracts the difference of the average sentiment (also the percentage of positive words minus the negative words) of all news on the same week. News volume is the division of the number of Bitcoin news and the

number of total news on the same week. But for SCMP data sample, due to the lack of total number of news, I only standardize the nominal number of Bitcoin news.

5.2.2 Google SVI

Google allows users to access data on relative search volume with Google Trends for a certain keyword. The relative search volume is scaled between 0 and 100. 0 indicates the lowest search interest whereas 100 indicates the maximum search interest in a specified time frame. R provides the package 'gtrendsR' for querying Google Trends data. With the API 'gtrends', I first retrieve daily search volume with keyword 'bitcoin' and geography 'CN', 'UK' and the default geography of 'Global' respectively. Because this data is indexed relative to the search volume of a month, it needs to be rescaled to create a time series of daily search index across years. To rescale the data, I download the monthly SVI indexed relative to a year's search volume directly from Google Trends website[2]. Finally, I calculate adjustment factors by dividing the sum of daily index for each month.

$$Adjustment\ Factor_{y,m} = \frac{Yearly\ Index_y}{\sum_{d=1}^{30,31} Daily\ Index_{y,m,d}}$$

$$Adjusted\ Daily\ Index_t = Adjustment\ Factor_{y,m} * Daily\ Index_t$$

- $Yearly\ Index_y$ denotes the monthly index for a specific year downloaded directly from Google Trends web portal. The year ranges from 2013 to 2019.

- $Adjustment\ Factor_{y,m}$ is the adjustment factor for each month in a particular year. For example, $Adjustment\ Factor_{2014,1}$ is the adjustment factor for January in 2014.

- $Adjusted\ Daily\ Index_t$ is the final adjusted daily search index.

- $Daily\ Index_t$ is the monthly adjusted daily index retrieved using gTrends API.

Alternatively, one can obtain the weekly indices directly using gTrends APIs. But because Google does not return weekly index for longer than three months period, the resulting weekly data still has to be fetched in several batches and rescaled. Regression models using SVI are added in my analysis to cross validate my conclusion.

---

[2] https://trends.google.com/trends/explore?date=2013-01-01%202019-07-01&q=bitcoin

The dummy variable *country* is added to identify the origin of a search index for cross-sectional analysis. Regression analysis using SVI complement the news-based models in investigating the existence of continental difference between investors' attention.

Comparing to SVI, news sentiment and news volume are both indirect measures of investor attentions. Direct measures are often considered superior since they do not suffer the same issues indirect measures do, e.g., derived inaccuracies of prior measurement or too restrictive assumptions. SVI is easy to acquire. But news contents require much more complicated processing prior to data analysis. SVI is also superior to news-based measures time vice. News events are less frequent comparing to SVI which is available on a daily basis.

5.3 Control variables

VIX, gold, FTSE 100 and Shanghai A-share from the traditional markets are used as control variables to proxy the macroeconomic situations which may impact Bitcoin pricing.

5.3.1 VIX

The Chicago Board of Exchange Market Volatility Index (VIX) is a key market risk indicator reflecting market sentiment and investor expectation. According to Thomas (Thomas, 2015), an increased VIX usually suggests elevated market uncertainty and leads to 'flights to safety'. The negative correlation between the VIX and safer assets, e.g., treasury bonds and gold, reveals investors' hedging strategy of moving away from risky assets into safe haven assets. Bouri, et al., (Bouri, et al., 2017) conclude the hedging ability of Bitcoin against equities and commodities. Their findings imply certain similarities between VIX and Bitcoin volatility.

An earlier study from MacDonell (MacDonell, 2014) concludes a negative correlation between VIX and Bitcoin price, suggesting that investors speculate on Bitcoin when volatility in traditional markets is low. In a more recent study, Estrada (Estrada, 2017) found no statistical significance of Granger-causality between Bitcoin price and VIX, but instead found bidirectional Granger-causality between Bitcoin trading volume and VIX.

5.3.2 Gold

Gold has been used as a store of wealth and a medium of exchange throughout history (Goodman, 1956). But since 1971 the decoupling of the US dollar from gold, a system of fiat currencies has been adopted globally (Kento, 2019). Nowadays, investors generally

invest in gold to shield wealth from economic uncertainty. The demand for gold increased drastically after the Global Financial Crisis in 2008 (Biakowski, et al., 2015).

Gold and cryptocurrencies have many similarities. Dyhrberg finds Bitcoin reacts to similar variables as gold does including exchange rates but with a higher frequency (Dyhrberg, 2015).

- Both have little utility or intrinsic value. Besides jewellery and coin fabrication, gold has only limited use in technology and medicine while the demand has been declining (WorldGoldCouncil, 2019).

- Both are not backed by any central authorities, hence are relatively insulated from macroeconomic downturns.

- Both have limited supply. Bitcoin's supply is capped to 21 million in total and predetermined to be reached by year 2140.

5.3.3 FTSE 100 Index

The Financial Times Stock Exchange 100 Index (FTSE 100 Index) is a share index of the 100 largest UK companies listed on the London Stock Exchange. FTSE comprises mostly international companies headquartered in Europe. Hence, it provides an overview of the state of the economy in Europe.

The GBP denominated FTSE price is used in the regression analysis. This is because the exchange rate from the GBP to the US dollar may carry certain North American bias towards Bitcoin and bring in noise to the data. Dyhrberg concludes that Bitcoin reacts positively to increases in the FTSE Index, but negatively to USD/GBP exchange rate (Dyhrberg, 2015), using daily data. I intend to focus mainly on the sentiment analysis, hence, excluding the exchange rate influence for now.

5.3.4 Shanghai A-share

Shanghai A-share tracks the daily price performance of all A-shares listed on the Shanghai Stock Exchange. Despite the clamping down on digital currency exchanges and banning the trading of Bitcoin in 2017, China, as one of the biggest economies in Asia, is still an active player in the cryptocurrency market. Bouoiyour et al. reveal a positive short run dependence of Bitcoin on the Shanghai market index (Bouoiyour & Selmi, 2015). Hence, Shanghai A-share may act as a potential source of Bitcoin price volatility. For

similar reasons as FTSE 100 index, the YUAN denominated Shanghai A-share price is used in my regressions.

5.4 Multiple linear regressions

The ordinary least-squares (OLS) regression is used to test the effect of investor sentiment and attention on Bitcoin return and prices. The preliminary models of the regressions for log return and log price are defined as below:

$$lnReturn = \mu + \beta_1 abnPos + \beta_2 newsVol + \beta_3 lnFtse + \beta_4 lnGold + \beta_5 lnVix + \beta_6 lnSse$$

$$lnPrice = \mu + \beta_1 abnPos + \beta_2 newsVol + \beta_3 ftse + \beta_4 gold + \beta_5 vix + \beta_6 sse$$

$$lnPrice = \mu + \beta_1 wsvi + \beta_2 ftse + \beta_3 gold + \beta_4 vix + \beta_5 sse$$

Appendix 2, Table 11 shows the pair-wise Pearson's correlation between (ln) Bitcoin return, news-based attention measures and (ln) returns of other control variables, measured at weekly frequency. The weekly log return of Bitcoin has very low correlations with all the other variables. Among the control variables, the return of FTSE has negative 70% correlation to VIX and 30% to SSE return, which may cause collinearity concerns in the OLS result.

Table 12 shows the correlation between variables log return, positive sentiment, negative sentiment, and the other control variables at monthly frequency. The correlation between monthly log return of Bitcoin and the other variables increase considerably comparing to its weekly frequency. Monthly log return of Bitcoin is negatively correlated with VIX, gold and SSE.

Table 13 shows the pair-wise Pearson's correlations among (ln) Bitcoin price, news-based attention measures and other variables of interest, measured at weekly frequency. In general, the correlations between log Bitcoin price and news volume are at about 60% for both FT and SCMP. The correlation between news volume and weekly SVI is very high at about 90% and 80% for FT and SCMP respectively, implying the resemblance of news volume to SVI as a measure of investor attention. Comparing to news volume, sentiment is much less correlated to price, at about 20% from FT dataset, and only 10% from SCMP datasets. VIF tests are conducted in section 5.6 to formally address the collinearity concerns for all regression models.

The correlation between sentiment (measured by percentage of sentimental words *abnPos*) and investor attention (measured by news volume *newsVol*) is fairly low relative to other variables, at only 4% and -10% from FT and SCMP datasets respectively as shown in Table 13. As discussed by Da, et al., "attention is a necessary condition for generating sentiment, increased investor attention...will likely lead to stronger sentiment.". Da et al. also use Loughran's approach to measure sentiment. But they found even lower correlation of 1.4% and 2.3% between news sentiment and attention measures proxied through SVI (Da, et al., 2011). The lower correlations in their findings are likely due to the discrepancy between the investor attention captured by SVI and the investor attention indirectly proxied through news sentiment. SVI certainly captures investors' attention but are these investors the same group of people paying attention to the news? In other words, the very low correlation may be partly due to the inconsistency between the target groups.

In this thesis I use news sentiment and news volume originated from the same news articles to measure investors' sentiment and attention to mitigate the mismatching issue. This likely explains the higher correlations between sentiment and attention in my findings. Interestingly the correlation between sentiment and news volume is even higher for SCMP samples, around -10%, suggesting as the news volume increases, the sentiment of these news become more negative. This is the opposite to FT samples with a positive relation between volume and sentiment.

Intuitively increased attention is believed to be accompanied by attenuated sentiment. But as shown in Figure 5, sentiment does not become too extreme as news volume increases. Observations cluster around zero in the range of minus two to plus two, indicating not too extreme sentiment, even when the news volume become very high. The relation between sentiment and attention is still unclear and is a very intriguing topic to investigate.

Figure 5   Scatter plot of news sentiment against news volume with fitted regression lines

The solid straight line is the fitted OLS regression line. The solid red curve line specifies the Loess estimate of the mean function of the vertical axis variable (relative news volume, *newsVol*) given the horizontal axis variable (news sentiment, *abnPos*) and the two dash-dot red lines are the Loess estimate of variability. Loess is a non-parametric approach that fits multiple regressions in local neighbourhood. *newsVol* and *abnPos* are defined in Appendix 1.
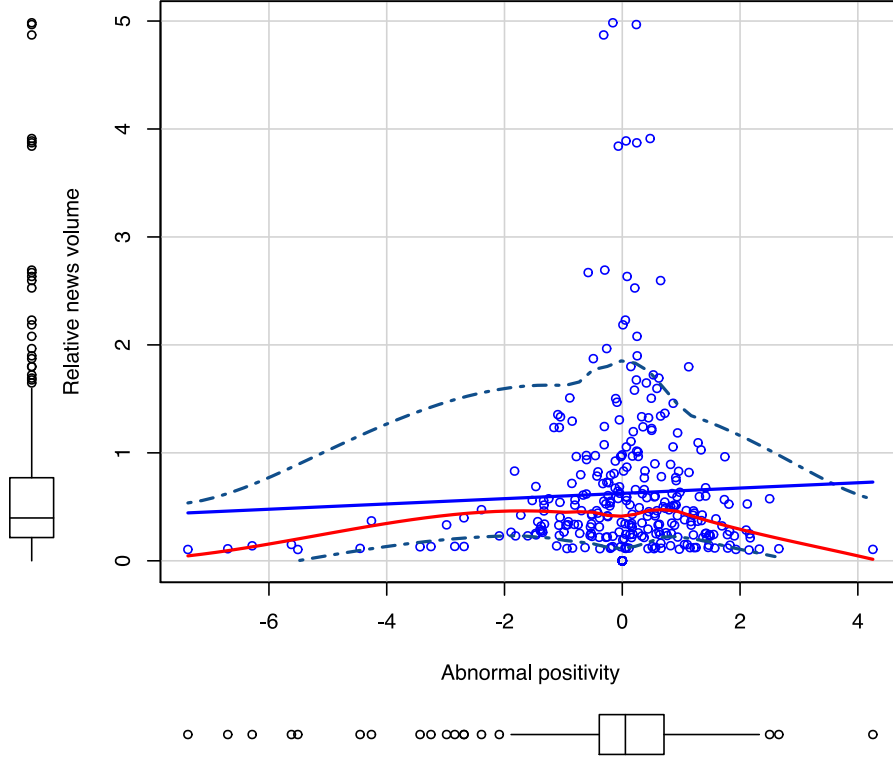
5.5   Results

5.5.1   *Bitcoin returns*

Table 3 shows the OLS regression results, where the dependent variable is always the weekly natural log return of Bitcoin. Neither news volume nor sentiment is significant in explaining the weekly return. The variation of the Bitcoin returns is mostly unexplained by the regression models with a very small R-squared. Furthermore, the main results show that there is no significant relationship between returns on Bitcoin and global proxies of traditional asset classes. These results are consistent with the empirical results of some studies that cryptocurrencies may be useful for diversification purposes, being uncorrelated with other asset classes (Bianchi, 2018). Dyhrberg also concluded the hedging ability of Bitcoin against FTSE index. The return on Bitcoin is not affected by changes in the assets in FTSE index in average (Dyhrberg, 2015) (Dyhrberg, 2016).

Considering the protentional influence of news sentiment impact on Bitcoin, I decided to restore the sentiment measure back to its two original components: the percent of positive words and the percent of negative words. Table 4 and Table 5 show the OLS regression results, where the dependent variable is always the natural log return of Bitcoin and the independent variables are the change of positive and negative sentiment and news volume at weekly and monthly frequencies respectively. The negative and positive sentiment is still insignificant at weekly interval as shown in Table 6. But for monthly data, the positive sentiment becomes significant at 5% level for all five models from FT datasets as well as the cross-sectional regressions. No significance is observed from SCMP samples.

In Table 7, both the positive and negative sentiment are significant and have positive impact on Bitcoin returns. This result empirically supports the hypothesis proposed by Florian et al. that positive news is more likely to exert a positive influence, because they attract new users and affirm existing users to stay invested (Florian, et al., 2014). Relative to positive sentiment, negative sentiment is less significant but still has positive impact on Bitcoin monthly return. As the common notion dictates: "There is no such thing as bad publicity". Negative news also attract attention, enhancing the desirability of Bitcoin.

Once again, the dummy variable *source* is insignificant in the cross-sectional models in both Table 6 and Table 7. Change of news volume is not significant in any models, neither does the global proxies of traditional asset classes at monthly frequency. Neither the interaction terms between dummy and sentiment nor between dummy and change of news volume is significant.

Table 3 **Regressions for weekly log return and sentiment**

This table regresses weekly Bitcoin return on sentiment measures and a few traditional assets for the FT and SCMP datasets respectively. The dependent variable is the weekly log return of Bitcoin. The independent variables are news sentiment measures as defined in Appendix 1. The rest are control variables, weekly log return of VIX (*lnVix*), Gold (*lnGold*), FTSE 100 (*lnFtse*) and Shanghai A-share (*lnSse*). The sample period is from 22nd April 2013 to 30th June 2019. Model 1, 2, 3 are from FT samples. Model 4, 5 and 6 are from SCMP samples. All the independent and control variables are standardized. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

| | *Dependent variable: lnReturn (weekly)* | | | | | | | | |
| | FT | | | SCMP | | | Cross-sectional | | |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
abnPos	0.002		0.002	0.001		0.0003	0.001		0.001
	(0.007)		(0.007)	(0.007)		(0.007)	(0.005)		(0.005)
newsVol		0.001	0.001		-0.003	-0.003		-0.001	-0.001
		(0.007)	(0.007)		(0.007)	(0.007)		(0.005)	(0.005)
lnVix	-0.001	-0.0002	-0.001	-0.001	-0.004	-0.004	0.0001	-0.0002	-0.001
	(0.069)	(0.069)	(0.069)	(0.069)	(0.069)	(0.069)	(0.049)	(0.049)	(0.049)
lnGold	0.002	-0.004	-0.002	-0.001	0.011	0.011	-0.029	-0.024	-0.025
	(0.408)	(0.409)	(0.410)	(0.408)	(0.408)	(0.409)	(0.295)	(0.296)	(0.296)
lnFtse	0.119	0.130	0.120	0.128	0.102	0.103	0.106	0.106	0.103
	(0.576)	(0.575)	(0.577)	(0.575)	(0.577)	(0.578)	(0.416)	(0.417)	(0.417)
lnSse							0.049	0.047	0.048
							(0.200)	(0.200)	(0.200)
source(SCMP)							-0.00000	0.00000	0.00000
							(0.010)	(0.010)	(0.010)
Constant	0.009	0.009	0.009	0.009	0.009	0.009	0.008	0.008	0.008
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Observations	323	323	323	323	323	323	634	634	634
R^2	0.001	0.0004	0.001	0.0003	0.001	0.001	0.001	0.0005	0.001
Adjusted R^2	-0.012	-0.012	-0.015	-0.012	-0.012	-0.015	-0.009	-0.009	-0.011
Residual Std. Error	0.126 (df = 318)	0.126 (df = 318)	0.127 (df = 317)	0.126 (df = 318)	0.126 (df = 318)	0.126 (df = 317)	0.127 (df = 627)	0.127 (df = 627)	0.127 (df = 626)
F Statistic	0.043 (df = 4; 318)	0.030 (df = 4; 318)	0.038 (df = 5; 317)	0.028 (df = 4; 318)	0.080 (df = 4; 318)	0.064 (df = 5; 317)	0.056 (df = 6; 627)	0.051 (df = 6; 627)	0.053 (df = 7; 626)

Note: *p**p***p<0.01

Table 4 **Regressions for weekly return and negativity and positivity**

This table regresses Bitcoin log return on sentiment measures and indices of traditional assets for the FT and SCMP datasets respectively. Model 1-5 are for FT datasets. Model 6-10 are from SCMP datasets. The dependent variable is the natural log of weekly Bitcoin return. The independent variables are percentage change of weekly news sentiment measures (*pos, neg*) and the percentage change of news volume $\Delta(newsVol)$. The rest are control variables, weekly log return of VIX (*lnVix*), Gold (*lnGold*), FTSE 100 (*lnFtse*). The sample period is from 22nd April 2013 to 30th June 2019. All definition of the variables can be found in Appendix 1. All the independent and control variables are standardized. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

	Dependent variable: lnReturn (weekly)									
	FT					SCMP				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
pos	-0.011		-0.010		-0.010	-0.010		-0.010		-0.009
	(0.013)		(0.013)		(0.013)	(0.013)		(0.013)		(0.014)
neg		-0.006	-0.005		-0.006		-0.001	-0.001		-0.0005
		(0.005)	(0.005)		(0.005)		(0.004)	(0.004)		(0.005)
$\Delta(newsVol)$				0.001	0.002				-0.003	-0.002
				(0.007)	(0.007)				(0.007)	(0.008)
lnVix	0.00001	-0.003	-0.002	-0.0002	-0.002	-0.009	-0.009	-0.010	-0.010	-0.010
	(0.069)	(0.069)	(0.069)	(0.069)	(0.069)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)
lnGold	0.011	0.026	0.033	-0.004	0.024	0.003	-0.026	0.002	-0.015	0.005
	(0.408)	(0.408)	(0.408)	(0.409)	(0.410)	(0.420)	(0.418)	(0.420)	(0.419)	(0.421)
lnFtse	0.159	0.117	0.145	0.130	0.150					
	(0.575)	(0.574)	(0.575)	(0.575)	(0.576)					
lnSse						0.063	0.058	0.063	0.048	0.058
						(0.279)	(0.279)	(0.280)	(0.280)	(0.281)
Constant	0.019	0.022	0.030*	0.009	0.031*	0.015	0.011	0.017	0.008	0.016
	(0.014)	(0.014)	(0.018)	(0.007)	(0.018)	(0.012)	(0.012)	(0.014)	(0.007)	(0.015)
Observations	323	323	323	323	323	317	317	317	317	317
R²	0.002	0.004	0.006	0.0004	0.006	0.002	0.001	0.002	0.001	0.002
Adjusted R²	-0.010	-0.009	-0.010	-0.012	-0.013	-0.011	-0.012	-0.014	-0.012	-0.017
Residual Std. Error	0.126 (df = 318)	0.126 (df = 318)	0.126 (df = 317)	0.126 (df = 318)	0.126 (df = 316)	0.127 (df = 312)	0.127 (df = 312)	0.127 (df = 311)	0.127 (df = 312)	0.128 (df = 310)
F Statistic	0.198 (df = 4; 318)	0.312 (df = 4; 318)	0.359 (df = 5; 318)	0.030 (df = 4; 317)	0.318 (df = 6; 318)	0.169 (df = 4; 316)	0.047 (df = 4; 312)	0.139 (df = 5; 312)	0.076 (df = 4; 311)	0.123 (df = 6; 310)

Note: *p**p***p<0.01

Table 5 **Regressions for monthly return and negativity and positivity**

This table regresses Bitcoin return on sentiment measures and indices of traditional assets for the FT and SCMP datasets respectively. Model 1-5 are for FT datasets. Model 6-10 are from SCMP datasets. The dependent variable is the natural log of monthly Bitcoin return. The independent variables are percentage change of monthly news sentiment measures (*pos, neg*) and the percentage change of news volume $\Delta(newsVol)$. The rest are control variables, monthly log return of VIX (*lnVix*), Gold (*lnGold*), FTSE 100 (*lnFtse*). The sample period is from 22nd April 2013 to 30th June 2019. All definition of the variables can be found in Appendix 1. All the independent and control variables are standardized. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

	Dependent variable: lnReturn (monthly)									
	FT					SCMP				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
pos	0.120**		0.138**		0.138**	0.078		0.100		0.097
	(0.056)		(0.056)		(0.056)	(0.059)		(0.061)		(0.060)
neg		0.076	0.097*		0.096*		0.040	0.059		0.064
		(0.052)	(0.051)		(0.051)		(0.041)	(0.042)		(0.042)
$\Delta(newsVol)$				-0.037	-0.038				-0.088	-0.094
				(0.070)	(0.067)				(0.081)	(0.081)
lnVix	-0.099	-0.074	-0.130	-0.055	-0.130	-0.119	-0.095	-0.131	-0.084	-0.123
	(0.250)	(0.254)	(0.246)	(0.256)	(0.247)	(0.185)	(0.185)	(0.184)	(0.185)	(0.184)
lnGold	-0.052	-0.088*	-0.065	-0.075*	-0.064	-0.063	-0.071	-0.058	-0.079*	-0.064
	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)
lnFtse	-0.264	-0.001	-0.262	-0.001	-0.234					
	(1.718)	(1.744)	(1.687)	(1.767)	(1.696)					
lnSse						-0.483	-0.431	-0.557	-0.497	-0.665
						(0.720)	(0.722)	(0.717)	(0.725)	(0.721)
Constant	0.059	0.044	0.049	0.053	0.050	0.044	0.057	0.046	0.049	0.042
	(0.042)	(0.043)	(0.041)	(0.043)	(0.042)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
Observations	74	74	74	74	74	74	74	74	74	74
R^2	0.105	0.074	0.150	0.050	0.154	0.074	0.063	0.099	0.066	0.117
Adjusted R^2	0.053	0.021	0.087	-0.005	0.078	0.020	0.009	0.033	0.012	0.038
Residual Std. Error	0.358 (df = 69)	0.364 (df = 69)	0.351 (df = 68)	0.369 (df = 69)	0.353 (df = 67)	0.364 (df = 69)	0.366 (df = 69)	0.362 (df = 68)	0.366 (df = 69)	0.361 (df = 67)
F Statistic	2.022 (df = 4; 69)	1.382 (df = 4; 69)	2.393** (df = 5; 68)	0.907 (df = 4; 69)	2.028* (df = 6; 67)	1.371 (df = 4; 69)	1.159 (df = 4; 69)	1.495 (df = 5; 68)	1.223 (df = 4; 69)	1.479 (df = 6; 67)

Note: $^{*}p^{**}p^{***}p<0.01$

Table 6 **Cross-sectional regressions for weekly return and negativity and positivity with interaction terms**

This table regresses Bitcoin log return on sentiment measures and indices of traditional assets for the panel data of both FT and SCMP datasets. The dependent variable is the natural log of weekly Bitcoin log return. The independent variables are percentage change of weekly news sentiment measures (*pos, neg*) and the percentage change of news volume Δ(*newsVol*). The rest are control variables, weekly log return of VIX (*lnVix*), Gold (*lnGold*), FTSE 100 (*lnFtse*) and Shanghai A-share (*lnSse*). Dummy variable (*source*) is added to indicate the publisher of the news and also interacted with positivity and negativity. The sample period is from 22nd April 2013 to 30th June 2019. All definition of the variables can be found in Appendix 1. All the independent and control variables are standardized. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

	Dependent variable:					
	lnReturn					
	(1)	(2)	(3)	(4)	(5)	(6)
pos	-0.009		-0.008	-0.009		-0.009
	(0.010)		(0.010)	(0.009)		(0.009)
neg		-0.004	-0.003	-0.003		-0.003
		(0.004)	(0.003)	(0.004)		(0.003)
Δ(*newsVol*)					0.001	0.002
					(0.007)	(0.007)
lnVix	0.001	-0.001	-0.0005	-0.0005	-0.001	-0.0004
	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)
lnGold	-0.007	-0.021	-0.004	-0.005	-0.025	-0.009
	(0.296)	(0.295)	(0.296)	(0.296)	(0.296)	(0.296)
lnFtse	0.132	0.104	0.125	0.123	0.100	0.124
	(0.416)	(0.416)	(0.416)	(0.416)	(0.417)	(0.418)
lnSse	0.055	0.053	0.058	0.058	0.046	0.058
	(0.200)	(0.200)	(0.200)	(0.200)	(0.200)	(0.200)
pos:source(SCMP)	-0.002		-0.002			
	(0.010)		(0.010)			
neg:source(SCMP)		0.001		0.001		
		(0.004)		(0.004)		
newsVol:source(SCMP)					-0.004	-0.003
					(0.010)	(0.010)
Constant	0.017*	0.015*	0.021*	0.021*	0.008	0.021*
	(0.009)	(0.009)	(0.011)	(0.011)	(0.005)	(0.011)
Observations	634	634	634	634	634	634
R^2	0.002	0.002	0.003	0.003	0.001	0.003
Adjusted R^2	-0.007	-0.008	-0.008	-0.008	-0.009	-0.009
Residual Std. Error	0.127 (df = 627)	0.127 (df = 627)	0.127 (df = 626)	0.127 (df = 626)	0.127 (df = 627)	0.127 (df = 625)
F Statistic	0.251 (df = 6; 627)	0.198 (df = 6; 627)	0.294 (df = 7; 626)	0.295 (df = 7; 626)	0.081 (df = 6; 627)	0.266 (df = 8; 625)

Note: *p**p***p<0.01

Table 7 Cross-sectional results for monthly return and sentiment with interaction term

This table regresses Bitcoin return on sentiment measures and indices of traditional assets with interaction term between sentiment and *source* dummy for the panel data of both FT and SCMP datasets. The dependent variable is the natural log of monthly Bitcoin return. The independent variables are percentage change of monthly news sentiment measures (*pos, neg*), the percentage change of news volume Δ(*newsVol*), the interaction term is between sentiment measures and *source* dummy. The rest are control variables, monthly log return of VIX (*lnVix*), Gold (*lnGold*), FTSE 100 (*lnFtse*) and Shanghai A-share (*lnSse*). The sample period is from 22nd April 2013 to 30th June 2019. All definition of the variables can be found in Appendix 1. All the independent and control variables are standardized. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

	Dependent variable: lnReturn (monthly)					
	(1)	(2)	(3)	(4)	(5)	(6)
pos	0.145***		0.161***	0.143***		0.142***
	(0.055)		(0.055)	(0.040)		(0.040)
neg		0.047	0.069**	0.074		0.072**
		(0.052)	(0.032)	(0.051)		(0.032)
Δ(*newsVol*)					-0.032	-0.031
					(0.071)	(0.068)
vix	-0.192	-0.167	-0.201	-0.204	-0.175	-0.214
	(0.174)	(0.179)	(0.172)	(0.172)	(0.180)	(0.173)
gold	0.839	0.590	0.758	0.789	0.507	0.616
	(0.785)	(0.803)	(0.776)	(0.774)	(0.826)	(0.797)
sse	-0.651	-0.522	-0.630	-0.625	-0.555	-0.635
	(0.510)	(0.526)	(0.503)	(0.507)	(0.527)	(0.506)
ftse	-0.305	-0.239	-0.260	-0.278	-0.386	-0.392
	(1.227)	(1.264)	(1.211)	(1.215)	(1.277)	(1.223)
pos:source(SCMP)	-0.047		-0.038			
	(0.079)		(0.078)			
neg:source(SCMP)		-0.006		-0.008		
		(0.067)		(0.064)		
Δ(*newsVol*):source(SCMP)					-0.029	-0.033
					(0.109)	(0.104)
Constant	0.052*	0.053*	0.049*	0.047	0.053*	0.046
	(0.030)	(0.031)	(0.029)	(0.029)	(0.031)	(0.029)
Observations	148	148	148	148	148	148
R²	0.082	0.029	0.112	0.110	0.022	0.115
Adjusted R²	0.043	-0.012	0.067	0.066	-0.019	0.065
Residual Std. Error	0.359 (df = 141)	0.369 (df = 141)	0.354 (df = 140)	0.354 (df = 140)	0.370 (df = 141)	0.354 (df = 139)
F Statistic	2.101* (df = 6; 141)	0.714 (df = 6; 141)	2.512** (df = 7; 140)	2.476** (df = 7; 140)	0.534 (df = 6; 141)	2.268** (df = 8; 139)

Note: *p**p***p<0.01

5.5.2 Bitcoin price

Appendix 3, Table 18 shows the OLS regression results, where the dependent variable is always natural log of weekly Bitcoin prices (*lnPrice*). Model 1, 2 and 3 are the regression results for FT dataset. Model 4, 5, and 6 are for SCMP dataset. Only difference is that model 1-3 use FTSE 100 as a control variable while model 4-6 use Shanghai A-share. Model 7-9 are the cross-sectional tests with the dummy variable *source* indicating the origin of a news observation.

In all models, *lnPrice* is positively related to the news-based attention measures (*abnPos* and *newsVol*). Controlling for other traditional assets, the coefficients associated with *abnPos* and *newsVol* remain strongly significant with the expected signs in almost all models, except 5. This shows that sentiment and attention measures both have significant explanatory power over Bitcoin price at weekly frequency. The test results reveal that the impact of new volume is economically and statistically greater than the impact of news sentiment.

In model 1, one standard deviation increases of news volume increases Bitcoin price by about 50% ($e^{0.404} - 1$, because the price is transformed into natural logarithms in the regression). In model 2, one standard deviation increase of sentiment (*abnPos*) leads to 12% increase in Bitcoin price. In the cross-sectional model 3, one standard deviation increase of news volume leads to 50% increase in price, but sentiment only leads to around 13% increase. The difference is even larger for SCMP datasets, where news volume leads to 1.4 times price increase in model 4, while sentiment becomes insignificant in model 5. In model 6, *newsVol* causes 1.5 times price increase but sentiment only contributes to around 24%. The positive effect of sentiment and attention appears to be consistent with Barber and Odean's findings for stock market that investors are more likely to buy, rather than sell, stocks that are in the news, leading to a price pressure (Barber & Odean, 2008).

5.5.3 Continental differences

The dummy variable (*source*) is not significant in neither the regressions for Bitcoin return nor those of Bitcoin price. For price-based regressions, sentiment and news volume remain positive and significant, but it does not matter whether the news is from FT or SCMP. Thus, the regression results fail to conclude that the sentiment of investors from Asia and from Europe have different impact on Bitcoin pricing.

According to this result, I conjecture two possibilities:

- Bitcoin topic is so international that all news sources uphold a relatively uniformed view without much continental discrimination;

- The text mining algorithm is too primitive to detect the subtleties and varieties in the contents.

To further understand how much news from SCMP and FT resemble each other, I revisit the original raw news data. I first tokenize each news to extract all vocabularies, and then calculate the overlapping rates of words as defined below:

$$overlapRate_i = \frac{numOverlapWords_i}{numTotalWords_i}$$

- $numOverlapWords_i$ is the number of words which are present in both $news_i$ (from SCMP) and $news_j$ (from FT).

- $numTotalWords_i$ is the total number of words in $news_i$ (from SCMP).

- Calculate the overlap rate from each news from the SCMP dataset against all news from the FT dataset and keep the largest overlap rate as the final result.

- Repeat the same process for each news from the FT dataset.

Figure 6   **Histogram of vocabulary overlapping rate**

Red is for the FT samples. Blue is for the SCMP samples. The X-axis is the overlap rates defined in section 5.5.3. The Y-axis is the number of Bitcoin news. The sampling period is from 22nd April 2013 to 30th June.
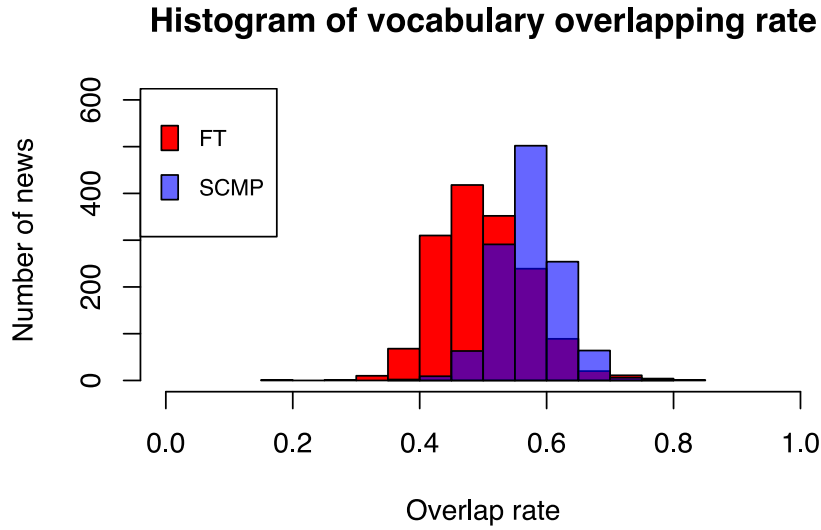
Figure 6 shows the histogram of vocabulary overlapping rates for the SCMP and the FT datasets. The overlapping rates cluster around 50% for the FT samples and 60% for the SCMP samples, suggesting the levels of content similarities from these two sources towards Bitcoin related topics. This may be one factor contributing to the statistical insignificance of continental difference between sentiments from SCMP and from FT. However, it is worth noting that my algorithm is only conducted at the word level, which does not account for context nor capture semantic similarities. More sophisticated algorithms such as machine learning approach likely provide results with higher precision. Researches about text mining of similarities is beyond the scope of this thesis.

5.5.4 *SVI*

Before rushing into any conclusions about the continental differences between investors' sentiment, I further conduct cross-sectional OLS regression analysis using the alternative attention proxy, SVI. Appendix 5, Table 22 presents the results regressed against Bitcoin price from the same day, next day, two days, one week and ten days after. Table 26 show the regression results using SVI from China and UK respectively as independent variables at weekly frequency. Both the daily and the weekly SVI (*wsvi*) is significant at 1% level in all models, revealing the high explanatory power of SVI to Bitcoin prices. But the dummy variable *country* is insignificant in all models. In Table

22, the interaction term between daily SVI and dummy variable *country* is significant at 5% and 1% level, but the dummy variable *country* is not significant by itself.

Using the alternative attention measure, SVI, I still fail to conclude that investors from different countries contribute to Bitcoin pricing differently. However, the SVI data for China may not be the best measure of Chinese investors' attention since Google moved away from mainland China to Hong Kong in 2010. Search data from Chinese search services probably would provide more information. The SVI from UK should be representative and informative, but it is still insignificant. Obviously, despite being highly correlated with Bitcoin price, SVI does not provide enough information to differentiate the regional impact of investors' sentiment.

Among the control variables, FTSE is consistently significant in all the models and has a large positive impact on Bitcoin price, with one standard deviation increase more than doubling the Bitcoin price. Shanghai A-share on the other hand has negative impact on Bitcoin price. In the daily models, the Bitcoin price rises as the VIX index rises. The positive effects of the VIX on Bitcoin prices seems to suggest that when volatility in traditional markets is high, investors seek out Bitcoin to shelter their wealth. Apparently, this conclusion is different from MacDonell's research which concluded an inverse relationship between Bitcoin and VIX values. When the VIX is low, investors push money towards Bitcoins (MacDonell, 2014). The VIX reached two significant peaks in February and December of 2018 respectively. Bitcoin however behaved inversely relative to the VIX on both occasions. MacDonell's research data was collected before my sampling period. The relation between the VIX and Bitcoin has been changing as time evolves, which likely causes the discrepancy between my results from MacDonell's. Considering the bumpy trail of Bitcoin development, adding a dummy to indicate different time period would be sensible. Gold is significant and negatively related to Bitcoin price in the daily models, which is expected as to a safe haven asset.

5.6   Model diagnosis

5.6.1   *Multiple collinearity*

The existence of collinearity destabilises the model and inflates the regression results. Formal tests are conducted after the models are established, using variance inflation factor (VIF). VIFs can explain one variable in terms of the linear combination of all the other variables. A VIF value less than or equals one suggests no collinearity; a value

between one and five means moderate collinearity; for variables which have a value higher than five should not be used directly in the regression.

Table 14 and Table 16 in Appendix 3 show the VIF results for monthly log return regressions. Table 19 shows the VIF results for the log price models regressed against sentiment measures. Appendix 5, Table 24 and Table 27 show the VIF results for the log price models against daily and weekly SVI. All VIF factors are less than 5 but slightly over 1, implying very moderate to none collinearity.

5.6.2 Heteroscedasticity

Another important assumption in OLS is that the variance of the error terms should be consistent as regressor values change, known as homoscedasticity. But such constraining assumption is difficult to adhere to due to the extreme movements of Bitcoin prices. I first conduct Breusch-Pagan test for all the models and conclude the existence of heteroskedasticity. Then, I estimate the heteroskedasticity consistent (HC) variance covariance matrix for the parameters. Table 15 and Table 17 are the HC-corrected p-value for monthly log return models. The significance levels even increased for the sentiment measure in model 1, 3 and 5 in Table 15 and as well as model 1, 3, 4 and 6 in Table 17. Table 20 shows the HC-corrected p-value for the log price and sentiment-based regressions. Table 28 shows the HC-corrected p-values for log price and weekly SVI models. The HC corrected p-values are slightly different from homoskedasticity-only models, but the significance levels remain unchanged. Table 25 shows the HC-corrected significance levels for Bitcoin price and daily SVI regressions. The p-value of the interaction term between SVI and country dummy becomes insignificant after the HC correction, but the significance level of other variables remains the same. Hence, the conclusions drawn from the regression models are robust.

6   GRANGER CAUSALITY ANALYSIS

In this section, I conduct Granger causality tests for the attention measures versus Bitcoin returns and prices. It is however worth noting that Granger causality does not prove actual causation, only that the two values are related by some phenomenon. There has been much criticism of Granger causality testing in the econometrics literature. Roberts and Nord found that the functional form of the time series affected the sensitivity of both Granger's and Sims' tests. Data that had undergone logarithmic transformation showed no sign of causality while the untransformed data yielded significant results (Roberts & Nord, 1985). Varying the functional form of the variables may lead to inconsistent results. Hence, to improve the robustness and reliability of the inferences, I conduct Granger tests over different functional forms of the variables in different combinations. All variables are first order integrated to make the time series stationary, based on the results of unit root tests.

6.1   Bitcoin return versus SVI and news-based measures

Firstly, I attempt to determine whether statistically SVI provides more information about future returns of Bitcoin than past values of Bitcoin return alone and vice versa at weekly and monthly frequencies. The lags chosen are 1-day, 2-day, 3-day and then 1-week, 1-month and 3-month. The first section in 0 show that daily SVI Granger causes Bitcoin return at lag 1, and vice versa for both SVI and log transformed SVI. For weekly frequency, log return consistently Granger causes SVI at lag 2 but not the other way around.

The lead-lag effect is much weaker between news-based measures and log return. No Granger-causality between sentiment and log return, and news volume leads log return at lag 52 for FT dataset. For SCMP dataset, news volume leads log return at lag 2 and log return leads sentiment at lag 52. It is thus difficult to draw any reliable conclusions from the results between sentiment and log return.

Table 8 **Granger test for SVI and Bitcoin return at daily and weekly frequencies**

All variables are first order integrated. The data is taken from 22nd April 2013 to 30th June 2019. The daily *lnReturn* is the natural log of Bitcoin opening price divided by closing price on the same day. *lnSVI* is the natural log of SVI. The first section is based on daily data. The second section is for weekly data. SVI is the worldwide SVI. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

Daily SVI and daily Bitcoin log return						
Null hypothesis	**1-lag**	**2-lag**	**3-lag**	**7-lag**	**30-lag**	**100-lag**
SVI does not Granger-cause lnReturn	4.77**	2.07	2.74**	2.19**	1.86***	1.78***
lnReturn does not Granger-cause SVI	3.34*	1.15	1.28	0.98	1.72***	0.97
lnSVI does not Granger-cause lnReturn	15.22***	1.43	1.8	1.55	1.91***	1.99***
lnReturn does not Granger-cause lnSVI	22.36***	8.93***	7.86***	4.22***	2.21***	1.18
Weekly SVI and weekly Bitcoin log return						
Null hypothesis	**1-lag**	**2-lag**	**3-lag**	**4-lag**	**52-lag**	**78-lag**
SVI does not Granger-cause lnReturn	0.7	1.51	2.6*	1.9	1.7***	1.45**
lnReturn does not Granger-cause SVI	0.44	4.86***	3.57**	2.8**	1.21	0.82
lnSVI does not Granger-cause lnReturn	0.08	0.42	2.68**	1.78	1.35*	1.01
lnReturn does not Granger-cause lnSVI	5.09**	3.63**	2.55*	2.07*	1.07	0.81
FT Sentiment and weekly Bitcoin log return						
Null hypothesis	**1-lag**	**2-lag**	**3-lag**	**4-lag**	**52-lag**	**78-lag**
lnReturn does not Granger-cause newsVol	0.08	0.79	0.64	0.99	0.89	1.01
newsVol does not Granger-cause lnReturn	0.54	0.29	1.37	1.24	1.72***	1.56**
lnReturn does not Granger-cause abnPos	0.37	0.33	0.21	0.37	0.96	1.04
abnPos does not Granger-cause lnReturn	0.03	1.53	1.12	1.13	0.88	0.67
SCMP Sentiment and weekly Bitcoin log return						
Null hypothesis	**1-lag**	**2-lag**	**3-lag**	**4-lag**	**52-lag**	**78-lag**
lnReturn does not Granger-cause newsVol	0.44	0.5	0.24	0.42	0.9	0.8
newsVol does not Granger-cause lnReturn	0.51	4.67**	2.81**	2.5**	1.35*	1.52**
lnReturn does not Granger-cause abnPos	0.6	0.41	0.99	1.04	1.67***	1.24
abnPos does not Granger-cause lnReturn	0	0.06	0.06	0.19	0.93	0.82

6.2 Bitcoin price versus SVI

Hypothetically investor attention would be aroused when Bitcoin price fluctuates dramatically; then attenuated investor attention further enhance the bumpy trail of

Bitcoin price. More specifically, investor attention leads to price increase, which empirically support the attention-induced price pressure hypothesis (Da, et al., 2011).

Table 29 shows the Granger causality between daily SVI and Bitcoin price. The result changes after SVI and price are log transformed. Therefore, it is inconclusive whether there exists any Granger causality between SVI and Bitcoin price at daily frequency. The lead-lag effect is much more consistent for the weekly data. For all the different combinations of functional transforms, there exists bidirectional Granger causality between SVI and Bitcoin price at lag 1 at least 10% significance level and the effect last for at least two weeks, suggesting the mutually enhancing effect between SVI and Bitcoin price.

6.3 SVI versus news measures

Next I investigate the Granger causality between SVI and the news-based attention measures at both the shorter-term lags of 1,2, 3, 4 weeks and longer-term lags of 1 year and 1.5 year. In both Table 30 and Table 31, the first four rows show the Granger test results between SVI and news volume for FT and SCMP datasets respectively. In weekly frequency, both SVI and log SVI lead news volume at lag 1. From the other direction, the result is inconsistent for FT dataset. Hence, I fail to conclude the existence of Granger causality from news volume to SVI. For SCMP dataset there is no Granger causality from news volume to SVI. The lead-lag effect from SVI to news volume may imply that investors may start paying attention to Bitcoin in anticipation of future events. But news contents are usually about events already happened. This is consistent with the finding of Da et al. from the stock market that log SVI lead news-related variables after major news and high turnover during which SVI spikes (Da, et al., 2011). The result suggests that SVI captures investor attention more promptly than news-based measures.

The fifth to eighth rows in Table 30 shows the unidirectional causality from bitcoin price to news volume significant at least 5% level at lag 1, suggesting that price fluctuation is followed up by the media in a week but not the other way around. For the SCMP dataset news volume starts to lead price at lag 2 and persists for a long time (see Table 31). The last two sections in Table 30 and Table 31 test the interaction between sentiment, attention and price. There is no consistent lead-lag effect between sentiment and news volume, neither between Bitcoin price nor sentiment from either direction.

7 CONCLUSION AND FUTURE WORK

According to my regression analysis, the impact of investor attention and sentiment is positive and very significant to Bitcoin price formation. Moreover, the investor attention (proxied through news volume) has considerably higher impact on Bitcoin prices than investor sentiment (proxied through news sentiment), which empirically supports the attention-induced price pressure hypothesis.

The continental differences of investor opinions between Asia and Europe (proxied through news published by SCMP and FT) is not observed. Further regression analysis using the alternative attention proxy SVI based on search data from China and UK also fails to conclude the existence of such continental discrimination. It is however still too premature to conclude that there is no such difference. Future research may consider using search data provided by Asian search services and news articles from more than one Asian providers in different languages to proxy Asian investor attention and sentiment.

The return of Bitcoin responds positively to both positive and negative news at monthly frequency. The positive sentiment has almost twice the impact on Bitcoin monthly return as much as negative sentiment does. The positive news is more likely to exert a positive influence. Negative news, on the other hand, has not caused significant price correction as expected. Once again, the origin of the news is insignificant. The continental difference of investor opinions is not observed at a monthly frequency. Change of news volume is not significant, neither are the global proxies of traditional asset classes including the monthly log return of VIX, gold, FTSE and SSE, at explaining the monthly return of Bitcoin.

There exists bidirectional Granger causality between Bitcoin price and SVI at lag 1 at least 10% significance level and the effect lasts for at least two weeks. The Granger causality is unidirectional from Bitcoin price to news volume at lag 1 for the FT dataset. For the SCMP dataset, news volume leads Bitcoin price at lag 2. For a weekly frequency, the Google SVI leads the news-based attention measure (news volume) at lag 1, implying that investors may start paying attention to Bitcoin in anticipation of future events. News reports then follow up the eye-catching events with some delay.

SVI appears to capture investor attention more promptly than news-based measures. The interaction between SVI and news-based attention measures is only superficially studied. Search data such as SVI is easy to acquire and simple to interpret. News articles

on the other hand, carry more comprehensive information and influence people's decision making in a more complicated way. It would be interesting to understand more thoroughly the interaction between search data proxied investor attention and the news-proxied investor attention. Future research may investigate whether search data is superior or complementary to news-based measures and in what way. No Granger causality is concluded between news sentiment and news volume, neither between Bitcoin price nor sentiment.

The sampling period is from 2013 to 2019 during which Bitcoin has experienced dramatic fluctuations. For future work, adding a time dummy to distinguish the different time periods would be more reasonable and likely uncover more insights. Some variables may respond differently at different time period. Endogenous factors such as Bitcoin transaction volume, hash rate, network volume, number of users shall be considered. Supply and demand are likely important factors at determining Bitcoin prices. More news data from different continents could be collected to better measure investor sentiment.

REFERENCES

Florian, G., Kai, Z. & Martin, H., 2014. *Bitcoin - Asset or Currency? Revealing Users' Hidden Intentions,* s.l.: Available at SSRN: https://ssrn.com/abstract=2425247.

Barberis, N., Shleifer, A. & Vishny, R. W., 1998. A model of investor sentiment. *Journal of Financial Economics,* 49(3), pp. 307-343 .

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

McCombs, M. & Reynolds, A., 2002. News influence on our pictures of the world. In: 2, ed. *Media Effects: Advances in Theory and Research.* s.l.:Lawrence Erlbaum Associates, pp. 1-18.

Ahern, K. R. & Sosyura, D., 2013. Rumor Has It: Sensationalism in Financial Media. *Review of Financial Studies,* 28(7), pp. 2050-2093.

Karlgren, J. et al., 2012. *Usefulness of Sentiment Analysis.* Barcelona, ECIR'12 Proceedings of the 34th European conference on Advances in Information Retrieval.

Merton, R. C., 1987. A simple model of capital market equilibrium with incomplete information. *Journal of Finance,* Volume 42, pp. 483-510.

Barber, B. M. & Odean, T., 2008. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies,* 21(2), pp. 785-818.

Fang, L. & Peress, J., 2009. Media Coverage and the Cross-section of Stock Returns. *Journal of Finance,* 64(5), pp. 2023-2052.

MacDonell, A., 2014. *Popping the Bitcoin Bubble: An application of log-periodic power law modeling to digital currency,* s.l.: University of Notre Dame.

Dyhrberg, A. H., 2015. Bitcoin, gold and the dollar - A GARCH volatility analysis. *Finance Research Letters,* Volume 16, pp. 85-92.

Bouri, E., Gupta, R., Tiwari, A. K. & RouBaud, D., 2017. Does Bitcoin hedge global uncertainty? Evidence from wavelet-based quantile-in-quantile regressions. *Finance Research Letters,* Volume 23, pp. 87-95.

Da, Z., Engelberg, J. & Gao, P., 2011. In Search of Attention. *Journal of Finance,* 66(5), pp. 1461-1499.

Cheah, E.-T. & Friy, J., 2015. Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics Letters,* Volume 130, pp. 32-36 .

Abraham, J., Higdon, D., Nelson, J. & Ibarra, J., 2018. Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis. *SMU Data Science Review,* 1(3).

Grinberg, R., 2012. Bitcoin: An Innovative Alternative Digital Currency. *Hastings Science and Technology Law Journal,* 4(1), p. 159.

Ramiah, V., Xu, X. & Moosa, I. A., 2015. Neoclassical finance, behavioral finance and noise traders: A review and assessment of the literature. *International Review of Financial Analysis,* Volume 41, pp. 89-100 .

Barberis, N. & Thaler, R., 2003. Chapter 18 A survey of behavioral finance. *Handbook of the Economics of Finance,* Volume 1, pp. 1053-1128 .

De Long, J. B., Shleifer, A., Summers, L. H. & Waldmann, R. J., 1990. Noise Trader Risk in Financial Markets. *Journal of Political Economy,* 98(4), pp. 703-738.

Shefrin, H. & Statman, M., 1994. Behavioral Capital Asset Pricing Theory. *Journal of Financial and Quantitative Analysis,* 29(3), pp. 323-349.

Baker, M. & Wurgler, J., 2007. Investor sentiment in the stock market. *Journal of Economic Perspectives,* Volume 21, pp. 129-151.

Da, Z., Engelberg, J. & Gao, P., 2014. The Sum of All FEARS Investor Sentiment and Asset Prices. *Review of Financial Studies,* 28(1), pp. 1-32.

Solt, M. & Statman, M., 1988. How Useful is the Sentiment Index?. *Financial Analysts Journal,* 44(5), pp. 45-55.

Verma, R. & Verma, P., 2007. Noise trading and stock market volatility. *Journal of Multinational Financial Management,* 17(3), pp. 231-243 .

Singer, E. & Ye, C., 2013. The Use and Effects of Incentives in Surveys. *The Annals of the American Academy of Political and Social Science,* 645(1), pp. 112-141.

Maynard, D. & Funk, A., 2011. *Automatic detection of political opinions in Tweets.* Heraklion, ESWC'11 Proceedings of the 8th international conference on The Semantic Web.

Boiy, E. & Moens, M.-f., 2009. A Machine Learning Approach to Sentiment Analysis in Multilingual Web Texts. *Information Retrieval,* 12(5), pp. 526-558.

Aue, A. & Gamon, M., 2005. *Customizing sentiment classifiers to new domains: A case study.* Borovets, Proceedings of the International Conference on Recent Advances in Natural Language Processing.

Loughran, T. & McDonald, B., 2011. When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance, Forthcoming,* 66(1), pp. 35-65.

Zhang, X., Fuehres, H. & Gloor, P. A., 2011. Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear". *Procedia - Social and Behavioral Sciences,* Volume 26, pp. 55-62.

Kearney, C. & Liu, S., 2014. Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis,* Volume 33, pp. 171-185.

Ferguson, N. J. & Philip, D., 2015. Media Content and Stock Returns: The Predictive Power of Press. *Multinational Finance Journal,* 19(1), pp. 1-31.

Kelly, S. & Ahmad, K., 2018. Estimating the impact of domain-specific news sentiment on financial assets. *Knowledge-Based Systems,* 150(15), pp. 116-126 .

Smales, L. & Lucey, B., 2019. The influence of investor sentiment on the monetary policy announcement liquidity response in precious metal markets. *Journal of International Financial Markets, Institutions and Money,* Volume 60, pp. 19-38.

Flori, A., 2019. News and subjective beliefs: A Bayesian approach to Bitcoin investments. *Research in International Business and Finance,* Volume 50, p. 336–356.

Simon, H. A., 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics,* 69(1), pp. 99-118.

Kristoufek, L., 2013. BitCoin meets Google Trends and Wikipedia: quantifying the relationship between phenomena of the Internet era. *Scientific Reports,* Volume 3.

Nasekin, S. & Chen, C. Y., 2019. *Deep Learning-Based Cryptocurrency Sentiment Construction,* s.l.: SSRN: http://dx.doi.org/10.2139/ssrn.3310784.

Dyck, A., Volchkova, N. & Zingales, L., 2008. The Corporate Governance Role of the Media: Evidence from Russia. *Journal of Finance,* 63(3), pp. 1093-1135.

Baker, S. R., Bloom, N. & Davis, S. J., 2016. Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics,* 131(4), p. 1593–1636.

Thomas, J. M., 2015. Structural relationships and portfolio efficiency. *Journal of Portfolio Management,* 42(1), pp. 135-142.

Estrada, J. C. S., 2017. *Analyzing Bitcoin Price Volatility,* s.l.: University of California, Berkeley.

Chan, W. H., Le, M. & Wu, Y. W., 2019. Holding Bitcoin longer: The dynamic hedging abilities of Bitcoin. *The Quarterly Review of Economics and Finance,* Volume 71, pp. 107-113.

Trimborn, S. & Härdle, W. K., 2018. Crix an index for cryptocurrencies. *Journal of Empirical Finance,* Volume 49, pp. 107-122.

CRypto IndeX, 2019. *RIX*. [Online]
Available at: https://thecrix.de/
[Accessed 10 August 2019].

Karabulut, Y., 2012. *Can Facebook Predict Stock Market Activity?,* s.l.: Available at SSRN: https://ssrn.com/abstract=2017099.

Modigliani, F. & Miller, M. H., 1958. The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review,* 48(3), pp. 261-297.

Markowitz, H. M., 1952. Portfolio selection. *Journal of Finance,* Volume 7, pp. 77-91.

Sharpe, W. F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance,* Volume 19, p. 425–442.

Black, F. & Scholes, M., 1973. The Pricing of options and corporate liabilities. *Journal of Political Economy,* Volume 81, p. 637–665.

Selden, G. C., 1912. *Psychology of the stock market: Human impulses lead to speculative disasters*. New York: Ticker Publishing.

Kahneman, D. & Tversky, A., 1979. Prospect theory: An analysis of decisions under risk. *Econometrica,* 47(2), p. 313–327.

Tversky, A. & Kahneman, D., 1986. Rational Choice and the Framing of Decisions. *The Journal of Business,* 59(4), pp. S251-S278.

Scheufele, D. A. & Tewksbury, D., 2007. Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models. *Journal of Communication,* 57(1), pp. 9-20.

Preis, T., Moat, H. S. & Stanley, H. E., 2013. Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports,* Volume 3, p. 1684.

Mondria, J., Wu, T. & Zhang, Y., 2010. The determinants of international investment and attention allocation: Using internet search query data. *Journal of International Economics,* Volume 82, pp. 85-95.

Choi, H. & Varian, H., 2012. Predicting the Present with Google Trends. *The Economic Record,* Volume 88, p. 2–9 .

Chong, D. & Druckman, J. N., 2007. Framing Theory. *Annual Review of Political Science,* Volume 10.

Valencia, F., Gómez-Espinosa, A. & Valdés-Aguirre, B., 2019. Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. *Entropy,* 21(6), p. 589.

Bianchi, D., 2018. *Cryptocurrencies As an Asset Class? An Empirical Assessment,* s.l.: Available at SSRN: https://ssrn.com/abstract=3077685.

Roberts, D. L. & Nord, S., 1985. Causality tests and functional form sensitivity. *Applied Economics,* 71(1), pp. 135-141.

Kento, W., 2019. *Nixon Shock*. [Online]
Available at: https://www.investopedia.com/terms/n/nixon-shock.asp
[Accessed 10 Oct 2019].

Biakowski, J., Bohl, M., Stephan, P. & Wisniewski, T., 2015. The gold price in times of crisis. *International Review of Financial Analysis ,* Volume 41, pp. 329-339.

WorldGoldCouncil, 2019. *metalsdaily*. [Online]
Available at: https://www.metalsdaily.com/uploads/gdt-fy-2018.pdf
[Accessed 16 Oct 2019].

Dyhrberg, A. H., 2016. Hedging capabilities of bitcoin. Is it the virtual gold?. *Finance Research Letters,* Volume 16, p. 139–144.

Goodman, B., 1956. The price of gold and international liquidity. *Journal of Finance,* Volume 11, pp. 15-28.

Bouoiyour, J. & Selmi, R., 2015. What Does Bitcoin Look Like?. *Annals of Economics and Finance,* 16(2), pp. 449-492.

Lamon, C., Nielsen, E. & Redondo, E., 2017. *Cryptocurrency Price Prediction Using News and Social Media Sentiment*. s.l., s.n.

Kaminski, J. & Gloor, P., 2014. Nowcasting the bitcoin market with twitter signals. *Computing Research Repository,* Volume abs/1406.7577.

Chousa, J. R. P., Cabarcos, M. L. & Pico, A. M. P., 2016. Examining the influence of stock market variables on microblogging sentiment. *Journal of Business Research ,* Volume 69, p. 2087––2092.

Dong, F., Xu, Z. & Zhang, Y., 2019. *Bubbly Bitcoin,* s.l.: SSRN:
    https://ssrn.com/abstract=3290125.

Karalevicius, V., Degrande, N. & Weerdt, J. D., 2018. Using sentiment analysis to
    predict interday Bitcoin price movements. *Journal of Risk Finance,* 19(1), pp.
    56-75.

Garcia, D., Juan Tessone, C., Mavrodiev, P. & Perony, N., 2014. *The digital traces of
    bubbles: feedback cycles between socio-economic signals in the Bitcoin
    economy,* s.l.: ArXiv e-prints, available at:
    http://adsabs.harvard.edu/abs/2014arXiv1408.1494G.

Ponsford, M., 2015. A Comparative Analysis of Bitcoin and Other Decentralised Virtual
    Currencies: Legal Regulation in the People's Republic of China, Canada, and the
    United States. *Hong Kong Journal of Legal Studies,* Volume 9, p. 29.

Figà-Talamanca, G. & Patacca, M., 2018. Does Market Attention Affect Bitcoin Returns
    and Volatility?. *Decisions in Economics and Finance,* p. Available at SSRN:
    https://ssrn.com/abstract=3148018.

Bleher, J. & Dimpfl, T., 2018. Today I Got a Million, Tomorrow, I Don't Know: On the
    Predictability of Cryptocurrencies by Means of Google Search Volume. p.
    Available at SSRN: https://ssrn.com/abstract=3126324.

APPENDIX 1   VARIABLE DEFINITION

Variable	Definition

Weekly variables

lnPrice   natural logarithm of weekly average Bitcoin price (from Monday to Sunday) denominated in US dollars.

abnPos $abnPos_i = (btcPos_i - avgPos_i) - (btcNeg_i - avgNeg_i)$

- $btcPos_i$ is the average of the percentages of positive words for Bitcoin news published on week i.

- $btcNeg_i$ is the average of the percentages of negative words for Bitcoin news published on week i.

- $avgPos_i$ is the average of the percentages of positive words for all news published on week i.

- $avgNeg_i$ is the average of the percentages of negative words for Bitcoin news published on week i.

newsVol   The number of Bitcoin news published weekly divides the total number of news of the same week $\frac{btcNewsVol_i}{totalNewsVol_i}$.

- $btcNewsVol_i$ is the number of Bitcoin on week i.

- $totalNews_i$ is the number of all news published on week i. The final result is standardized. This calculation is only applied to FT dataset. SCMP news volume (number of news published each week) is used directly after standardization.

source   Dummy variable, 'FT' means the news is from Financial Times; 'SCMP' means the news is from South China Morning Post.

vix[3]   The weekly average VIX denominated in USD. Missing 27-May 2019 data.

gold[4]   The weekly average gold price in USD.

ftse[5]   The weekly average of FTSE 100 index in GBP.

---

[3] https://finance.yahoo.com/quote/%5EVIX/history/
[4] https://www.quandl.com/data/WGC/GOLD_DAILY_USD-Gold-Prices-Daily-Currency-USD
[5] https://finance.yahoo.com/quote/%5EFTSE%3FP%3DFTSE/history/

sse[6]	Weekly average price of Shanghai A-share in CNY.
country	Dummy variable to identify the origin of the search volume, 'CN' means the SVI is from China; 'UK' means the SVI is from Britain; 'Global' means the SVI is worldwide.
wsvi	The weekly (from Monday to Sunday) average of *svi*. See section 5.2.2 for the detailed derivation.
lnReturn	natural logarithm of weekly Bitcoin returns denominated in US dollars, $\ln\left(\frac{P_t}{P_{t-1}}\right)$, P_t is Sunday's price, P_{t-1} is Monday's price of the same week.
lnVix	natural logarithm return of weekly VIX denominated in USD, $\ln\left(\frac{P_t}{P_{t-1}}\right)$, P_t is the Friday price, P_{t-1} is from the Monday of the same week.
lnGold	natural logarithm return of weekly gold denominated in USD, $\ln\left(\frac{P_t}{P_{t-1}}\right)$, P_t is the Friday price, P_{t-1} is from the Monday of the same week.
lnFtse	natural logarithm return of weekly FTSE 100 denominated in GBP, $\ln\left(\frac{P_t}{P_{t-1}}\right)$, P_t is the price on Friday, P_{t-1} is from the Monday of the same week.
lnSse	natural logarithm return of weekly Shanghai A-share denominated in Yuan, $\ln\left(\frac{P_t}{P_{t-1}}\right)$, P_t is the price on Friday, P_{t-1} is from the Monday of the same week.

Monthly data

lnReturn	natural logarithm return of monthly Bitcoin price denominated in USD, $\ln\left(\frac{P_t}{P_{t-1}}\right)$, P_t is the of price of the last day of a month, P_{t-1} is from the first day of the same month.
pos	natural logarithm of monthly news sentiment, $\ln\left(\frac{pos_t}{pos_{t-1}}\right)$, pos_t is the percentage of positive words of the last sample in a month, pos_{t-1} is the first sample of the same month. If the percentage of positive words is zero for the first sample, use the closest previous day's data.
neg	natural logarithm of monthly news sentiment, $\ln\left(\frac{neg_t}{neg_{t-1}}\right)$, neg_t is the percentage of negative words of the last sample of a month, neg_{t-1} is the first sample of the same month. If the percentage of negative words is zero for the first sample, use the closest previous day's data.

---

[6] https://finance.yahoo.com/quote/000001.SS/history?p=000001.SS

$\Delta(newsVol)$ natural logarithm of monthly news volume, $\ln\left(\frac{newsVol_t}{newsVol_{t-1}}\right)$ is

- $newsVol_t$ the number of news on the last sample of a month

- $newsVol_{t-1}$ is the first sample of the same month. If the number of news is zero for the first sample, use the closest previous day's data.

$lnVix$ natural logarithm return of monthly VIX denominated in USD, $\ln\left(\frac{P_t}{P_{t-1}}\right)$, P_t is the of price of the last trading day of a month, P_{t-1} is from the first trading day of the same month.

$lnGold$ natural logarithm return of monthly Gold denominated in USD, $\ln\left(\frac{P_t}{P_{t-1}}\right)$, P_t is the of price of the last trading day of a month, P_{t-1} is from the first trading day of the same month.

$lnFtse$ natural logarithm return of monthly FTSE 100 denominated in GBP, $\ln\left(\frac{P_t}{P_{t-1}}\right)$, P_t is the of price of the last trading day of a month, P_{t-1} is from the first trading day of the same month.

$lnSse$ natural logarithm return of monthly Shanghai A-share denominated in Yuan, $\ln\left(\frac{P_t}{P_{t-1}}\right)$, P_t is the of price of the last trading day of a month, P_{t-1} is from the first trading day of the same month.

APPENDIX 2   DESCRIPTIVE STATISTICS

Table 9 **Descriptive statistics of weekly observations**

Descriptive statistics of Bitcoin price, VIX, gold, Shanghai A-share, FTSE 100, sentiment score (*abnPos*) and news volume for FT and SCMP datasets respectively, and Google SVI. All data are taken at weekly frequency. Google SVI is standardized. Observations are from 22nd April 2013 to 30th June 2019. Shanghai A-share misses six weeks data.

Weekly	N	Mean	St. Dev.	Min	Max
BTC price	323	2554.926	3421.620	78.864	17512.450
vix	323	14.929	3.800	9.340	31.846
gold	323	1255	73.43	1062	1469
sse	317	2916.196	589.193	1960.691	5127.879
ftse	323	6882.385	463.151	5647.700	7794.420
lnReturn	323	0.009	0.126	-0.664	1.141
lnVix	323	-0.019	0.143	-0.479	0.767
lnGold	323	-0.0005	0.018	0.076	0.04
lnSse	317	-0.0003	0.026	-0.130	0.092
lnFtse	323	0.001	0.017	-0.057	0.095
(FT) abnPos	293	1.471	1.379	-2.954	8.832
(FT) newsVol	323	5.195	6.169	0	38
(SCMP) abnPos	259	1.728	1.757	-2.309	11.060
(SCMP) newsVol	323	3.375	3.854	0	22
svi	323	0.000	1.000	-0.574	7.456

Table 10 **Descriptive statistics of monthly observations**

Monthly	N	Mean	St. Dev.	Min	Max
(FT) pos	74	-0.047	0.769	-1.725	1.679
(FT) neg	74	0.110	0.836	-1.908	2.503
(SCMP) pos	74	0.120	0.739	-2	2
(SCMP) neg	74	-0.074	1.040	-2.597	2.643
lnReturn	74	0.052	0.368	-2.143	1.327
lnVix	74	-0.0001	0.254	-0.706	0.838
lnGold	74	0	0.038	-0.084	0.116
lnSse	74	0.003	0.065	-0.237	0.204
lnFtse	74	0.002	0.037	-0.112	0.095

Table 11 **Correlation matrix for variables used in weekly log return regressions**

This table shows the pair-wise Pearson's correlations between variables at weekly frequency. The variables are defined in Appendix 1. The sample period is from 22nd April 2013 to 30th June 2019.

	lnReturn	lnVix	lnGold	lnSse	lnFtse	(FT) newsVol	(FT) abnPos	(SCMP) newsVol
lnVix	-0.01							
lnGold	-0.02	0.1						
lnSse	0.01	-0.2	-0.04					
lnFtse	0.02	-0.7	0.02	0.3				
(FT) newsVol	0.01	-0.01	-0.01	-0.02	-0.01			
(FT) abnPos	0.02	-0.02	0.04	0	0.1	0.04		
(SCMP) newsVol	-0.03	-0.05	-0.1	-0.1	-0.04	0.7	-0.02	
(SCMP) abnPos	0.01	0.05	0.05	-0.01	-0.03	-0.04	0.1	-0.1

Table 12 **Correlation matrix for variables used in monthly log return regressions**

This table shows the correlations among variables at monthly frequency. The variables are defined in Appendix 1. The sample period is from 22nd April 2013 to 30th June 2019.

	(FT) pos	(FT) neg	(SCMP) pos	(SCMP) neg	lnReturn	lnVix	lnGold	lnSse
(FT) neg	-0.2							
(SCMP) pos	0.3	0.01						
(SCMP) neg	0.02	0.2	-0.2					
lnReturn	0.3	0.1	0.2	0.1				
lnVix	0.02	0.1	0.1	-0.01	-0.1			
lnGold	-0.2	0.2	-0.1	0.02	-0.1	0.1		
lnSse	0.04	-0.2	0.04	0.04	-0.1	-0.4	0.05	
lnFtse	0	-0.1	-0.01	-0.04	0.02	-0.7	0.2	0.4

Table 13 **Correlation matrix for variables used in log price regressions**

This table shows the pair-wise Pearson correlations between variables at weekly frequency. The variables are defined in Appendix 1. The sample period is from 22nd April 2013 to 30th June 2019.

	(FT) newsVol	(FT) abnPos	vix	gold	ftse	lnPrice	(SCMP) newsVol	(SCMP) abnPos	sse
(FT) abnPos	0.04								
vix	-0.1	0.01							
gold	0.17	-0.05	-0.2						
ftse	0.4	0.1	-0.5	0.3					
lnPrice	0.6	0.2	-0.1	0.2	0.8				
(SCMP) newsVol	0.7	-0.02	-0.05	0.2	0.5	0.6			
(SCMP) abnPos	-0.04	0.1	-0.1	0.01	0.05	0.1	-0.1		
sse	0.2	0.05	-0.1	-0.4	0.2	0.2	0.1	0.04	
wsvi	0.9	0.05	-0.2	0.2	0.5	0.6	0.8	0	0.2

APPENDIX 3   MODEL DIAGNOSIS FOR LOG RETURN BASED REGRESSIONS

Table 14 **VIF values of the monthly log return regression models**

This table shows the VIF values for the regression models presented in Table 5. Model 1-5 are from FT samples. Model 6-10 are from SCMP samples. All definition of the variables can be found in Appendix 1, under the section of monthly data.

	FT					SCMP				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
pos	1.07		1.1		1.1	1.04		1.12		1.12
neg		1.05	1.08		1.08		1	1.08		1.09
$\Delta(newsVol)$				1	1				1.04	1.06
lnVix	2.29	2.28	2.3	2.28	2.3	1.22	1.2	1.22	1.2	1.22
lnGold	1.12	1.09	1.14	1.05	1.15	1.06	1.03	1.07	1.05	1.09
lnFtse	2.25	2.24	2.25	2.24	2.25					
lnSse						1.2	1.19	1.2	1.2	1.22

Table 15 **Heteroskedasticity consistent p-value for monthly log return models**

This table shows the heteroskedasticity consistent significance levels for the regression models presented in Table 5. All definition of the variables can be found in Appendix 1.

	FT					SCMP				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
pos	0.03		0.01		0.01	0.2		0.15		0.19
neg		0.21	0.11		0.11		0.35	0.24		0.22
$\Delta(newsVol)$				0.62	0.59				0.4	0.41
lnVix	0.56	0.72	0.5	0.77	0.5	0.34	0.46	0.33	0.5	0.38
lnGold	0.38	0.22	0.32	0.25	0.32	0.32	0.26	0.33	0.25	0.32
lnFtse	0.85	1	0.85	1	0.87					
lnSse						0.31	0.37	0.3	0.27	0.22

Table 16 **VIF values of the monthly log return cross-sectional regression models**

This table shows the VIF values for the regression models presented in Table 7. *source* is the dummy variable, given the value of 'FT' or 'SCMP' indicating if the news was published by FT or SCMP. All definition of the variables can be found in Appendix 1, under the section of monthly data.

	(1)	(2)	(3)	(4)	(5)	(6)
pos	1.96		1.99	1.07		1.07
neg		2.61	1.07	2.67		1.08
$\Delta(newsVol)$					1.79	1.79
lnVix	2.21	2.21	2.21	2.22	2.22	2.23
lnGold	1.03	1.02	1.03	1.02	1.07	1.08
lnFtse	2.28	2.29	2.28	2.29	2.32	2.32
lnSse	1.24	1.25	1.24	1.25	1.24	1.25
pos:source	1.97		1.98			
neg:source		2.6		2.6		
$\Delta(newsVol):source$					1.84	1.85

Table 17 **Heteroskedasticity consistent p-value for cross-sectional monthly log return models**

This table shows the heteroskedasticity consistent significance levels for the regression models presented in Table 7. *source* is the dummy variable, given the value of 'FT' or 'SCMP' indicating if the news was published by FT or SCMP. All definition of the variables can be found in Appendix 1.

	(1)	(2)	(3)	(4)	(5)	(6)
pos	0.01		0.01	0		0
neg		0.29	0.06	0.11		0.05
$\Delta(newsVol)$					0.67	0.66
lnVix	0.22	0.32	0.21	0.2	0.29	0.19
lnGold	0.25	0.41	0.3	0.27	0.47	0.4
lnFtse	0.76	0.81	0.79	0.77	0.71	0.69
lnSse	0.06	0.13	0.09	0.1	0.1	0.08
pos:source	0.57		0.64			
neg:source		0.93		0.9		
$\Delta(newsVol):source$					0.82	0.8

APPENDIX 4   REGRESSIONS BASED ON WEEKLY LOG PRICE AND NEWS SENTIMENT

Table 18 **Weekly average Bitcoin price and sentiment**

This table regresses log Bitcoin price on sentiment measures and indices of traditional assets for the FT and SCMP datasets. The dependent variable is the natural log of weekly average Bitcoin price. The independent variables are abnormal positivity (*abnPos*) and the relative news volume (*newsVol*). The rest are control variables: weekly average indices of VIX (*vix*), Gold (*gold*), FTSE 100 (*ftse*) and Shanghai A-share (*sse*). The sample period is from 22nd April 2013 to 30th June 2019. Model 1, 2, 3 are from FT samples. Model 4, 5 and 6 are from SCMP samples. Model 7, 8, 9 are cross sectional regressions using both SCMP and FT datasets. *source* is the dummy variable, given the value of 'FT' or 'SCMP' indicating if the news was published by FT or SCMP. All definition of the variables can be found in Appendix 1. All independent and control variables are standardized. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

	Dependent variable: lnPrice								
	FT			SCMP			Cross-sectional		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
newsVol	0.404***		0.406***	0.854***		0.887***		0.415***	0.428***
	(0.044)		(0.044)	(0.061)		(0.061)		(0.032)	(0.031)
abnPos		0.111**	0.117***		0.069	0.202***	0.083***		0.120***
		(0.044)	(0.039)		(0.076)	(0.059)	(0.031)		(0.027)
gold	-0.019	0.018	-0.009	0.173**	0.425***	0.157**	-0.046	-0.097***	-0.095***
	(0.040)	(0.045)	(0.040)	(0.070)	(0.085)	(0.069)	(0.037)	(0.033)	(0.033)
vix	0.458***	0.529***	0.448***	-0.049	-0.027	-0.037	0.534***	0.441***	0.436***
	(0.046)	(0.050)	(0.045)	(0.062)	(0.079)	(0.061)	(0.036)	(0.033)	(0.032)
ftse	1.141***	1.339***	1.120***				1.395***	1.179***	1.160***
	(0.052)	(0.052)	(0.051)				(0.039)	(0.039)	(0.039)
sse				0.225***	0.393***	0.210***	-0.112***	-0.137***	-0.138***
				(0.068)	(0.084)	(0.067)	(0.037)	(0.033)	(0.033)
source(SCMP)							-0.0001	-0.002	-0.002
							(0.061)	(0.055)	(0.054)
Constant	6.905***	6.905***	6.905***	6.887***	6.890***	6.887***	6.899***	6.897***	6.898***
	(0.039)	(0.043)	(0.038)	(0.059)	(0.076)	(0.058)	(0.043)	(0.039)	(0.038)
Observations	323	323	323	317	317	317	634	634	634
R²	0.761	0.705	0.768	0.446	0.106	0.466	0.705	0.766	0.773
Adjusted R²	0.758	0.701	0.764	0.439	0.094	0.458	0.702	0.764	0.771
Residual Std. Error	0.695 (df = 318)	0.772 (df = 318)	0.686 (df = 317)	1.058 (df = 312)	1.345 (df = 312)	1.041 (df = 311)	0.771 (df = 627)	0.686 (df = 627)	0.676 (df = 626)
F Statistic	253.609*** (df = 4; 318)	190.015*** (df = 4; 318)	209.915*** (df = 5; 317)	62.919*** (df = 4; 312)	9.241*** (df = 4; 312)	54.388*** (df = 5; 311)	249.867*** (df = 6; 627)	342.345*** (df = 6; 627)	304.966*** (df = 7; 626)

Note: *p**p***p<0.01

Table 19 **VIF values of the log price regression models**

This table shows the VIF values for the regression models presented in Table 18. Model 1, 2, 3 are from FT samples. Model 4, 5 and 6 are from SCMP samples. Model 7, 8, 9 are cross sectional regressions using both SCMP and FT datasets. *source* is the dummy variable, given the value of 'FT' or 'SCMP' indicating if the news was published by FT or SCMP. All definition of the variables can be found in Appendix 1, under the section of weekly data.

	FT			SCMP			Cross-sectional		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
abnPos		1.02	1.02		1.01	1.03	1.01		1.02
newsVol	1.31		1.31	1.08		1.1		1.36	1.37
gold	1.09	1.09	1.09	1.39	1.3	1.4	1.48	1.5	1.5
vix	1.41	1.36	1.41	1.07	1.07	1.07	1.33	1.4	1.4
ftse	1.78	1.43	1.81				1.61	1.99	2.01
sse				1.3	1.26	1.3	1.47	1.47	1.47
source							1	1	1

Table 20 **Heteroskedasticity consistent p-value for log price and sentiment models**

This table shows the heteroskedasticity consistent significance levels for the regression models presented in Table 18. Model 1, 2, 3 are from FT samples. Model 4, 5 and 6 are from SCMP samples. Model 7, 8, 9 are cross sectional regressions using both SCMP and FT datasets. *source* is the dummy variable, given the value of 'FT' or 'SCMP' indicating if the news was published by FT or SCMP. All definition of the variables can be found in Appendix 1.

	FT			SCMP			Cross-sectional		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
abnPos		0.01	0.002		0.27	0.0002	0.005		0
newsVol	0		0	0		0		0	0
gold	0.52	0.86	0.67	0.04	0	0.06	0.11	0.001	0.002
vix	0	0	0	0.4	0.71	0.52	0	0	0
ftse	0	0	0				0	0	0
sse				0.004	0	0.006	0.003	0	0
source							1	0.97	0.97

Table 21 **Next-week average Bitcoin price and sentiment**

This table regresses log Bitcoin price on sentiment measures and indices of traditional assets for the FT and SCMP respectively. The dependent variable is the next week's average Bitcoin log price. The independent variables are news sentiment measure, abnormal positivity (*abnPos*) and the relative news volume (*newsVol*). The rest are control variables, weekly average indices of VIX (*vix*), Gold (*gold*), FTSE 100 (*ftse*) and Shanghai A-share (*sse*). The sample period is from 22nd April 2013 to 30th June 2019. Model 1, 2, 3 are from FT samples. Model 4, 5 and 6 are from SCMP samples. Model 7, 8, 9 are cross sectional regressions using both SCMP and FT datasets. *source* is the dummy variable, given the value of 'FT' or 'SCMP' indicating if the news was published by FT or SCMP. All definition of the variables can be found in Appendix 1. All the independent and control variables are standardized. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

	Dependent variable: lnPrice (next week)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
newsVol	0.396***		0.398***	0.849***		0.882***		0.407***	0.421***
	(0.045)		(0.044)	(0.062)		(0.061)		(0.032)	(0.031)
abnPos		0.107**	0.113***		0.073	0.206***	0.083***		0.119***
		(0.044)	(0.039)		(0.076)	(0.059)	(0.031)		(0.027)
gold	-0.017	0.019	-0.008	0.176**	0.427***	0.160**	-0.044	-0.095***	-0.092***
	(0.041)	(0.045)	(0.040)	(0.070)	(0.085)	(0.069)	(0.037)	(0.033)	(0.033)
vix	0.447***	0.517***	0.438***	-0.061	-0.039	-0.049	0.521***	0.430***	0.425***
	(0.046)	(0.050)	(0.046)	(0.062)	(0.079)	(0.061)	(0.036)	(0.033)	(0.032)
source(SCMP)							-0.0001	-0.002	-0.002
							(0.061)	(0.055)	(0.054)
ftse	1.145***	1.339***	1.125***				1.394***	1.182***	1.164***
	(0.052)	(0.052)	(0.052)				(0.039)	(0.039)	(0.039)
sse				0.227***	0.394***	0.211***	-0.111***	-0.135***	-0.137***
				(0.068)	(0.084)	(0.067)	(0.037)	(0.033)	(0.033)
Constant	6.919***	6.919***	6.919***	6.899***	6.903***	6.900***	6.911***	6.910***	6.910***
	(0.039)	(0.043)	(0.038)	(0.060)	(0.076)	(0.059)	(0.043)	(0.039)	(0.038)
Observations	323	323	323	317	317	317	634	634	634
R^2	0.759	0.705	0.765	0.444	0.108	0.465	0.706	0.765	0.772
Adjusted R^2	0.756	0.701	0.761	0.437	0.097	0.456	0.703	0.762	0.769
Residual Std. Error	0.699 (df = 318)	0.773 (df = 318)	0.691 (df = 317)	1.062 (df = 312)	1.345 (df = 312)	1.044 (df = 311)	0.770 (df = 627)	0.689 (df = 627)	0.680 (df = 626)
F Statistic	250.271*** (df = 4; 318)	189.989*** (df = 4; 318)	206.524*** (df = 5; 317)	62.304*** (df = 4; 312)	9.482*** (df = 4; 312)	53.995*** (df = 5; 311)	250.822*** (df = 6; 627)	339.243*** (df = 6; 627)	301.967*** (df = 7; 626)

Note: *p**p***p<0.01

APPENDIX 5   REGRESSIONS BASED ON SVI

Table 22 **Regression against Daily SVI**

This table regresses daily natural log Bitcoin price on Google SVI from China and UK respectively. The dependent variable is the natural log daily Bitcoin price (from the same day, next-day, 2-day, 6-day and 10-day). The rest are control variables, daily indices of VIX (*vix*), Gold (*gold*), FTSE 100 (*ftse*) and Shanghai A-share (*sse*). The sample period is from 1st April 2013 to 30th June 2019. *country* is the dummy variable, given the value of 'CN' or 'UK'. All definition of the variables can be found in Appendix 1. All the independent variables are standardized. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

	Dependent variable: lnPrice				
	0-day (1)	1-day (2)	2-day (3)	6-day (4)	10-day (5)
svi	0.436***	0.437***	0.436***	0.434***	0.435***
	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)
country (UK)	-0.024	-0.024	-0.024	-0.024	-0.024
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
vix	0.454***	0.453***	0.453***	0.447***	0.441***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
gold	-0.109***	-0.109***	-0.110***	-0.113***	-0.111***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
ftse	1.176***	1.176***	1.178***	1.181***	1.179***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
sse	-0.154***	-0.154***	-0.156***	-0.160***	-0.158***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
svi:country (UK)	-0.068***	-0.067***	-0.067***	-0.064**	-0.063**
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Constant	6.891***	6.893***	6.894***	6.903***	6.907***
	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)
Observations	2,894	2,894	2,894	2,894	2,894
R^2	0.754	0.755	0.756	0.758	0.758
Adjusted R^2	0.754	0.755	0.756	0.757	0.757
Residual Std. Error (df = 2886)	0.704	0.702	0.701	0.699	0.699
F Statistic (df = 7; 2886)	1,265.754***	1,271.722***	1,279.649***	1,289.337***	1,290.914***

Note: *p**p***p<0.01

Table 23 **Correlation matrix for the daily SVI models**

This table shows the correlations between variables used for the regression models presented in Table 22.

	svi (CN)	svi (UK)	gold	vix	ftse
svi (UK)	0.8				
gold	0.2	0.2			
vix	-0.1	-0.1	-0.2		
ftse	0.5	0.4	0.2	-0.5	
sse	0.2	0.1	-0.4	-0.05	0.2

Table 24 **VIF values of the daily SVI models**

This table shows the VIF values for the regression models presented in Table 22. *gold, vix, ftse, sse* are the daily closing prices of Gold, VIX, FTSE 100 and Shanghai A-share. *svi:country* is the interaction term of daily *svi* and its region (China or UK).

	(1)	(2)	(3)	(4)	(5)
svi	2.68	2.68	2.68	2.68	2.68
gold	1.44	1.44	1.44	1.44	1.44
vix	1.34	1.34	1.34	1.34	1.34
ftse	1.86	1.86	1.86	1.86	1.86
sse	1.45	1.45	1.45	1.45	1.45
country	1	1	1	1	1
svi:country	2.26	2.26	2.26	2.26	2.26

Table 25 **Heteroskedasticity consistent p-value for daily SVI models**

This table shows the heteroskedasticity consistent significance levels for the regression models presented in Table 22. *country* is the dummy variable, given the value of 'CN' or 'UK'

	(1)	(2)	(3)	(4)	(5)
svi	0	0	0	0	0
country (UK)	0.35	0.35	0.35	0.35	0.35
gold	0	0	0	0	0
vix	0	0	0	0	0
ftse	0	0	0	0	0
sse	0	0	0	0	0
svi:country	0.17	0.18	0.18	0.19	0.19

Table 26 **Regression results against Weekly SVI**

This table regresses logarithm Bitcoin price on Google SVI datasets from China and UK. The dependent variable is the natural logarithm same week average Bitcoin price for model 1 to 4, and next week average price for model 5 to 8. The rest are control variables, weekly average indices of VIX (*vix*), Gold (*gold*), FTSE 100 (*ftse*) and Shanghai A-share (*sse*). The sample period is from 22nd April 2013 to 30th June 2019. *country* is the dummy variable, given the value of 'CN' or 'UK'. All definition of the variables can be found in Appendix 1. All the independent variables are standardized. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

	Dependent variable:lnPrice							
	0-week				1-week			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
svi	1.012***	0.420***	0.486***	0.512***	1.012***	0.423***	0.482***	0.506***
	(0.059)	(0.045)	(0.032)	(0.042)	(0.059)	(0.045)	(0.032)	(0.042)
gold	0.070	-0.024	-0.115***	-0.116***	0.071	-0.023	-0.106***	-0.106***
	(0.065)	(0.040)	(0.032)	(0.032)	(0.065)	(0.040)	(0.032)	(0.032)
sse	0.005		-0.175***	-0.176***	0.006		-0.168***	-0.169***
	(0.065)		(0.032)	(0.032)	(0.065)		(0.032)	(0.032)
ftse		1.131***	1.134***	1.133***		1.130***	1.127***	1.125***
		(0.051)	(0.038)	(0.038)		(0.051)	(0.038)	(0.038)
vix	0.026	0.476***	0.456***	0.456***	0.013	0.463***	0.428***	0.428***
	(0.057)	(0.045)	(0.031)	(0.031)	(0.057)	(0.045)	(0.031)	(0.031)
svi:country (UK)				-0.050				-0.046
				(0.052)				(0.053)
country (UK)			-0.003	-0.003			-0.003	-0.003
			(0.052)	(0.052)			(0.053)	(0.053)
Constant	6.893***	6.905***	6.898***	6.898***	6.906***	6.919***	6.924***	6.924***
	(0.054)	(0.038)	(0.037)	(0.037)	(0.054)	(0.038)	(0.038)	(0.038)
Observations	317	323	634	634	317	323	634	634
R^2	0.535	0.765	0.783	0.784	0.536	0.766	0.779	0.779
Adjusted R^2	0.529	0.762	0.781	0.781	0.530	0.763	0.777	0.777
Residual Std. Error	0.970 (df = 312)	0.690 (df = 318)	0.661 (df = 627)	0.661 (df = 626)	0.970 (df = 312)	0.689 (df = 318)	0.668 (df = 627)	0.668 (df = 626)
F Statistic	89.776*** (df = 4; 312)	258.555*** (df = 4; 318)	377.725*** (df = 6; 627)	323.861*** (df = 7; 626)	90.044*** (df = 4; 312)	260.163*** (df = 4; 318)	368.313*** (df = 6; 627)	315.674*** (df = 7; 626)

Note: *p**p***p<0.01

Table 27 **VIF values of the weekly SVI regression models**

This table shows the VIF values for the regression models presented in Table 26. All definition of the variables can be found in Appendix 1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
wsvi	1.21	1.34	1.47	2.57	1.2	1.34	1.47	2.57
gold	1.44	1.09	1.51	1.51	1.44	1.09	1.51	1.51
vix	1.07	1.38	1.36	1.37	1.07	1.38	1.36	1.37
ftse		1.79	2.05	2.05		1.79	2.05	2.05
sse	1.42		1.48	1.49	1.42		1.47	1.49
svi:country				2.01				2.01
country			1	1			1	1

Table 28 **Heteroskedasticity consistent p-value for weekly SVI model**

This table shows the heteroskedasticity consistent significance levels for the regression models presented in Table 26. Model 1-4 are for the same week Bitcoin price. Model 5-8 are for the next week Bitcoin price. *country* is the dummy variable, given the value of 'CN' or 'UK'. All definition of the variables can be found in Appendix 1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
wsvi	0	0	0	0	0	0	0	0
country (UK)			0.96	0.96			0.96	0.96
gold	0.38	0.57	0	0	0.37	0.6	0.001	0.001
vix	0.68	0	0	0	0.83	0	0	0
ftse		0	0	0		0	0	0
sse	0.94		0	0	0.93		0	0
svi:country				0.57				0.61

APPENDIX 6   GRANGER CAUSALITY ANALYSIS

Table 29 **Granger test for SVI and Bitcoin price at daily and weekly frequency**

All variables are first order integrated. The data is taken from 22nd April 2013 to 30th June 2019. *lnPrice* is the natural logarithms of Bitcoin price. *lnSVI* is the natural logarithms of SVI. The first section is based on daily data. The second section are weekly average. SVI is the worldwide SVI. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

Daily SVI and Daily Bitcoin price						
NULL hypothesis	**1-lag**	**2-lag**	**3-lag**	**7-lag**	**30-lag**	**100-lag**
SVI does not Granger-cause price	1.38	29.78***	20.55***	25.56***	21.61***	16.04***
price does not Granger-cause SVI	5.26**	3.38**	0.82	10.4***	11.89***	7.57***
SVI does not Granger-cause lnPrice	3.43*	3.08**	2.13*	2.3**	2.06***	1.76***
lnPrice does not Granger-cause SVI	0.07	2.16	2.02	1.63	1.85***	0.99
lnSVI does not Granger-cause lnPrice	8.42***	3.93**	3.03**	2.29**	2.62***	1.98***
lnPrice does not Granger-cause lnSVI	1.2	12.85***	9.89***	5.12***	2.47***	1.27**
lnSVI does not Granger-cause price	0.12	1.03	0.74	2.03**	2.37***	1.73***
price does not Granger-cause lnSVI	0.43	1.41	1.2	1.63	1.42*	1.03

Weekly SVI and Weekly Bitcoin price						
NULL hypothesis	**1-lag**	**2-lag**	**3-lag**	**4-lag**	**52-lag**	**78-lag**
SVI does not Granger-cause price	29.52***	40.37***	28.33***	24.74***	14.63***	14.04***
price does not Granger-cause SVI	40.23***	21.73***	26.74***	19.93***	7.01***	9.98***
SVI does not Granger-cause lnPrice	3.62*	4.8***	3.48**	3.04**	2.02***	1.46**
lnPrice does not Granger-cause SVI	5.7**	2.83*	2.46*	1.98*	0.92	0.79
lnSVI does not Granger-cause lnPrice	3.67*	6.79***	4.9***	4.36***	1.44**	0.95
lnPrice does not Granger-cause lnSVI	3.83*	2.53*	1.78	1.55	0.97	0.75
lnSVI does not Granger-cause price	8.56***	8.41***	5.88***	4.59***	1.84***	1.13
price does not Granger-cause lnSVI	7.23***	4.23**	3.46**	2.8**	0.99	1.21

Table 30**Granger causality test for attention measures and Bitcoin price on FT dataset**

All variables are first order integrated and at weekly frequency from FT dataset. The data is taken from 22nd April 2013 to 30th June 2019. *lnPrice* is the natural logarithms of Bitcoin price. *lnSVI* is the natural logarithms of SVI. SVI is the worldwide SVI. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

SVI and FT news volume						
NULL hypothesis	**1-lag**	**2-lag**	**3-lag**	**4-lag**	**52-lag**	**78-lag**
SVI does not Granger-cause newsVol	7.71***	10.76***	16.23***	12.71***	2.02***	1.62**
newsVol does not Granger-cause SVI	9.82***	5.37***	3.62**	3.36**	1.82***	1.84***
lnSVI does not Granger-cause newsVol	2.81*	2.2	3.51**	2.55**	0.9	0.67
newsVol does not Granger-cause lnSVI	0.19	0.1	0.69	0.49	1.08	1.04
Bitcoin price and FT news volume						
NULL hypothesis	**1-lag**	**2-lag**	**3-lag**	**4-lag**	**52-lag**	**78-lag**
price does not Granger-cause newsVol	18.5***	14.06***	12.94***	12.33***	2.92***	2.6***
newsVol does not Granger-cause price	1.02	7.2***	4.48***	4.55***	1.88***	1.64**
lnPrice does not Granger-cause newsVol	3.85*	2.96*	2.37*	2.15*	1.11	1.09
newsVol does not Granger-cause lnPrice	0.29	1.88	1.59	1.35	1.26	1.44**
Bitcoin price and FT news sentiment						
NULL hypothesis	**1-lag**	**2-lag**	**3-lag**	**4-lag**	**52-lag**	**78-lag**
price does not Granger-cause abnPos	0.1	0.02	0.01	0.02	0.32	0.37
abnPos does not Granger-cause price	0.44	0.36	0.5	0.5	0.44	0.45
lnPrice does not Granger-cause abnPos	0.16	0.18	0.23	0.42	1.07	1.25
abnPos does not Granger-cause lnPrice	0	0.16	0.1	0.14	0.67	0.75
FT news volume and news sentiment						
NULL hypothesis	**1-lag**	**2-lag**	**3-lag**	**4-lag**	**52-lag**	**78-lag**
newsVol does not Granger-cause abnPos	1.36	0.5	0.52	0.75	0.33	0.6
abnPos does not Granger-cause newsVol	0.7	0.41	0.39	0.39	0.92	1.04

Table 31 **Granger causality test for attention measures and Bitcoin price on SCMP dataset**

All variables are first order integrated and at weekly frequency from SCMP dataset. The data is taken from 22nd April 2013 to 30th June 2019. *lnPrice* is the natural logarithms of Bitcoin price. *lnSVI* is the natural logarithms of SVI. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

SVI and SCMP news volume						
NULL hypothesis	**1-lag**	**2-lag**	**3-lag**	**4-lag**	**52-lag**	**78-lag**
SVI does not Granger-cause newsVol	4.53**	4.52**	3.24**	4.94***	3.62***	1.93***
newsVol does not Granger-cause SVI	0.4	1.96	2.73**	2.13*	1.86***	1.69***
lnSVI does not Granger-cause newsVol	5.13**	4.41**	4.5***	5.18***	1.58**	1.53**
newsVol does not Granger-cause lnSVI	0.08	0.31	0.28	0.81	1.07	1.11
Bitcoin price and SCMP news volume						
NULL hypothesis	**1-lag**	**2-lag**	**3-lag**	**4-lag**	**52-lag**	**78-lag**
price does not Granger-cause newsVol	0.14	2.3	1.14	1.55	2.58***	1.75***
newsVol does not Granger-cause price	1.32	11.29***	6.52***	5.49***	2.2***	2.32***
lnPrice does not Granger-cause newsVol	0	0.45	1.98	1.13	1.23	0.85
newsVol does not Granger-cause lnPrice	3.95**	4.1**	2.9**	2.59**	1.19	1.16
Bitcoin price and SCMP news sentiment						
NULL hypothesis	**1-lag**	**2-lag**	**3-lag**	**4-lag**	**52-lag**	**78-lag**
price does not Granger-cause abnPos	0	0.03	0.08	0.12	0.35	0.51
abnPos does not Granger-cause price	0.22	0.19	0.29	0.31	0.43	0.39
lnPrice does not Granger-cause abnPos	0.63	0.48	0.45	0.55	1.52**	1.66**
abnPos does not Granger-cause lnPrice	0.02	0.37	0.24	0.29	0.78	0.88
SCMP news volume and news sentiment						
NULL hypothesis	**1-lag**	**2-lag**	**3-lag**	**4-lag**	**52-lag**	**78-lag**
newsVol does not Granger-cause abnPos	3.34*	2.53*	1.96	1.66	0.56	0.91
abnPos does not Granger-cause newsVol	0.41	0.85	0.98	1.02	0.61	0.61

APPENDIX 7   TEN MOST NEGATIVE NEWS FROM FT

Publish date	Negativity	Content summary
2014-08-30	9.3	The Bitcoin entrepreneur Charlie Shrem reached a plea deal to resolve US charges in a scheme to sell over $1m of digital currency to users of illicit online marketplace Silk Road[7].
2015-08-01	9.1	Japanese police arrested Mark Karpelès, the head of Mt Gox[8].
2013-10-08	9.1	UK police arrested four men for alleged links to Silk Road[9].
2018-08-10	8.7	UK police warned of a rise in cryptocurrency fraud[10].
2019-06-05	8.5	The CEO of Longfin Corp was indicted on federal securities fraud charges for not engaging in any revenue-producing cryptocurrency transactions, nor use the blockchain to empower any solution[11].
2019-03-15	8.4	Japan court sentences Mark Karpelès, Mt Gox CEO[12].
2016-03-16	8.3	A Los Angeles hospital paid a bitcoin ransom equivalent of about $17,000 to retrieve its medical records after hackers attacked its network[13].
2013-07-23	8.1	SEC sued a Texas man for allegedly running a Ponzi scheme he promoted through a Bitcoin-denominated investment strategy[14].
2015-03-30	8.1	Two US agents investigating Silk Road were charged with stealing hundreds of thousands of dollars in bitcoin and other digital currencies[15].
2015-08-20	7.4	Mark Karpelès, Mt Gox CEO faced rearrest in bitcoin probe[16].

[7] https://www.ft.com/content/96396560-3031-11e4-9914-00144feabdc0
[8] https://www.ft.com/content/7c719a8c-3835-11e5-8613-07d16aad2152
[9] https://www.ft.com/content/ac25f088-2ff5-11e3-9eec-00144feab7de
[10] https://www.ft.com/content/f5583d68-9c9e-11e8-9702-5946bae86e6d
[11] https://www.ft.com/content/ee91aaa2-87ce-11e9-97ea-05ac2431f453
[12] https://www.ft.com/content/1e66c8ea-46c3-11e9-b168-96a37d002cd3
[13] https://www.ft.com/content/3b040e70-be8d-11e5-9fdb-87b8d15baec2
[14] https://www.ft.com/content/f5c0b9ea-f3cc-11e2-942f-00144feabdc0
[15] https://www.ft.com/content/81ed8e70-d710-11e4-97c3-00144feab7de
[16] https://www.ft.com/content/f4595a04-471b-11e5-af2f-4d6e0e5eda22

APPENDIX 8   TEN MOST NEGATIVE NEWS FROM SCMP

Date	Negativity	Content summary
2018-12-27	10.9	MtGox chief Mark Karpeles accused of pocketing bitcoin maintains innocence in Tokyo trial's closing arguments[17].
2014-09-24	10.5	The head of technology for Liberty Reserve, an alternative digital currency used in what authorities called a massive global fraud scheme, has pleaded guilty to criminal charges[18].
2017-06-28	10.4	A Florida man was sentenced to serve five and a half years in prison for operating an illegal bitcoin exchange suspected of laundering money for hackers[19].
2013-10-03	10.3	US law enforcement authorities shut down Silk Road and arrested its alleged owner[20].
2014-11-08	9.6	Police closed hundreds of online "dark" markets selling illegal drugs, weapons and services, arresting 17 people in a massive international operation against the Tor network[21].
2018-04-03	9.4	Alibaba sued a Dubai-based firm for using its trademarked name to raise more than US$3.5 million in cryptocurrency known as "Alibabacoins."[22]
2019-01-30	9.3	An insurance agent threatened to harm clients unless they paid him bitcoin has been jailed in Singapore[23].
2015-04-01	9.1	Two US agents investigating Silk Road were charged with stealing hundreds of thousands of dollars in bitcoin and other digital currencies[24].
2015-08-02	8.9	MtGox Bitcoin CEO faces new charges for embezzling $8.9 million in customer funds[25]
2018-02-09	8.5	Global cyberfraud gang's cofounder arresteed in Thailand[26]

[17] https://www.scmp.com/news/asia/east-asia/article/2179697/former-mtgox-chief-mark-karpeles-accused-pocketing-bitcoin
[18] https://www.scmp.com/news/world/article/1599722/digital-currency-executive-pleads-guilty-fraud-and-money-laundering?_escaped_fragment_=&edition=hong-kong
[19] https://www.scmp.com/news/world/united-states-canada/article/2100356/us-bitcoin-exchange-operator-jailed-laundering
[20] https://www.scmp.com/news/world/article/1323183/online-drug-marketplace-silk-road-shut-down-fbi-arrest-owner
[21] https://www.scmp.com/news/world/article/1634713/crackdown-darknet-results-17-arrests-and-closure-hundreds-sites-selling
[22] https://www.scmp.com/news/world/middle-east/article/2140012/alibaba-files-us-trademark-lawsuit-over-dubai-firms
[23] https://www.scmp.com/news/asia/southeast-asia/article/2184316/bitcoin-bandit-singaporean-insurance-agency-called-himself
[24] https://www.scmp.com/news/world/article/1752685/us-justice-department-charges-two-former-officials-stealing-bitcoins
[25] https://www.scmp.com/news/asia/east-asia/article/1845931/mtgox-bitcoin-ceo-faces-new-charges-embezzling-89-million
[26] https://www.scmp.com/news/asia/southeast-asia/article/2132742/global-cyberfraud-gangs-co-founder-arrested-thailand

APPENDIX 9   TEN MOST POSITIVE NEWS FROM FT

Date	Positivity	Content summary
2018-05-11	4.0	Nvidia revenues boosted by data centres and gamers[27].
2015-11-11	3.8	Osborne wants London to be global center for Fintech[28].
2018-12-04	3.8	Argo Blockchain said crypto enthusiasts are still clamoring for its coin mining service despite a market collapse[29].
2015-03-17	3.7	The whoosh of start-up value going from Europe to Palo Alto[30].
2014-11-05	3.4	Dining on slurry is not worth further discussion[31].
2018-01-03	3.3	Authers Note: Sehr optimistisch[32]
2018-12-04	3.2	Mediterranean EU countries make push on blockchain technology[33]
2013-06-30	3.2	Banking is heading towards its Spotify moment[34]
2018-03-05	3.2	Trustworthy data will transform the world[35]
2017-08-30	3.2	Estonian proposals of issuing its own cryptocurrency[36]

[27] https://www.ft.com/content/0c20ceec-5489-11e8-b24e-cad6aa67e23e
[28] https://www.ft.com/content/1f24a25e-886f-11e5-90de-f44762bf9896
[29] https://www.ft.com/content/88ff36c4-f795-11e8-8b7c-6fa24bd5409c
[30] https://www.ft.com/content/357cd0e8-c656-11e4-a13d-00144feab7de
[31] https://www.ft.com/content/047c68f2-638a-11e4-8a63-00144feabdc0
[32] https://www.ft.com/content/e437eaa4-f01b-11e7-b220-857e26d1aca4
[33] https://www.ft.com/content/95b57148-f7e1-11e8-af46-2022a0b02a6c
[34] https://www.ft.com/content/e1ae654a-c791-11e2-9c52-00144feab7de
[35] https://www.ft.com/content/d75f9cca-2052-11e8-a895-1ba1f72c2c11
[36] https://www.ft.com/content/32233b44-8ca1-11e7-a352-e46f43c5825d

APPENDIX 10 TEN MOST POSITIVE NEWS FROM SCMP

Date	Positivity	Content summary
2018-09-27	4.3	Bitmain-backed cryptocurrency unicorn Circle to launch US dollar 'stable coin' on Asian exchanges[37]
2018-02-30	3.8	Bitcoin pushes through US$11,000 threshold in fourth day of gains[38]
2019-02-14	3.6	Cryptocurrency 101: What is a stable coin?[39]
2017-09-22	3.2	Ant Financial remains optimistic on blockchain technology despite PBOC's cryptocurrencies crackdown[40]
2018-04-17	3.1	IMF chief says world must be even-handed in dealing with digital currencies' risks and innovations[41]
2017-10-19	3.0	Watch out: China's big ball of money may be headed into the stock market[42]
2017-04-11	3.0	Alibaba affiliate Ant Financial to accelerate blockchain initiatives[43]
2017-06-27	2.9	Large enterprises and start-ups focus on tech-driven solutions[44]
2017-11-09	2.8	Cheap and nasty money is what's keeping a zombie global economy afloat[45]
2017-09-08	2.8	Hong Kong regulators to push for greater clarity in Fintech market guidance to help companies develop[46]

[37] https://www.scmp.com/business/article/2165882/bitmain-backed-cryptocurrency-unicorn-circle-launch-us-dollar-stable-coin

[38] https://www.scmp.com/business/money/money-news/article/2133877/bitcoin-pushes-through-us11000-threshold-fourth-day-gains

[39] https://www.scmp.com/tech/blockchain/article/2185955/cryptocurrency-101-what-stable-coin

[40] https://www.scmp.com/tech/article/2112481/ant-financial-remains-optimistic-blockchain-technology-despite-pbocs

[41] https://www.scmp.com/business/banking-finance/article/2141983/world-must-be-even-handed-handling-digital-currencies-risks

[42] https://www.scmp.com/business/money/markets-investing/article/2116121/watch-out-chinas-big-ball-money-may-be-headed-stock

[43] https://www.scmp.com/tech/china-tech/article/2086449/alibaba-affiliate-ant-financial-accelerate-blockchain-initiatives

[44] https://www.scmp.com/special-reports/business/topics/hkex-special/article/2100174/large-enterprises-and-start-ups-focus

[45] https://www.scmp.com/business/companies/article/2119084/cheap-and-nasty-money-whats-keeping-zombie-global-economy-afloat

[46] https://www.scmp.com/tech/enterprises/article/2110239/hong-kong-regulators-push-greater-clarity-fintech-market-guidance

APPENDIX 11   THIRTY MOST FREQUENT TONAL WORDS OCCURRING IN ALL NEWS SAMPLES

Negative	Count	Positive	Count
AGAINST	1,169	GOOD	959
LAUNDERING	606	BEST	761
CRISIS	579	BETTER	662
LOST	522	ABLE	565
LATE	492	DESPITE	564
PROBLEM	486	INNOVATION	559
FRAUD	486	GREAT	518
CONCERNS	485	POPULAR	489
VOLATILITY	461	LEADING	421
QUESTION	401	STRONG	345
ILLEGAL	395	EASY	327
PROBLEMS	392	GREATER	324
THREAT	382	OPPORTUNITY	266
WARNED	380	GAINS	263
CUT	375	SUCCESS	256
QUESTIONS	365	GAIN	235
LOSSES	349	STABILITY	235
CRIMINAL	334	BOOM	234
DIFFICULT	329	IMPROVE	227
CLAIMS	328	SUCCESSFUL	225
BAD	312	EASIER	219
SHUT	308	ADVANTAGE	217
LACK	308	OPPORTUNITIES	211
CRIME	295	BENEFIT	194
INVESTIGATION	294	BOOST	191
STOLEN	282	EFFICIENT	187
CRIMINALS	281	HIGHEST	182
FORCE	278	EASILY	171
CHALLENGE	278	STABLE	166
FAILED	260	TRANSPARENCY	165