

Using sentiment analysis to predict interday Bitcoin price movements

Vytautas Karalevicius, Niels Degrande and Jochen De Weerd
KU Leuven, Leuven, Belgium

Abstract

Purpose – The purpose of this study is to measure the interaction between media sentiment and the Bitcoin price. Because some researchers argued that the Bitcoin value is also determined by perception of users and investors, this paper examines how.

Design/methodology/approach – The database of relative news articles as well as blog posts has been collected for the purpose of this research. Hence, each article has been given a sentiment score depending on the negative and positive words used in the article.

Findings – This paper has identified that interaction between media sentiment and the Bitcoin price exists, and that there is a tendency for investors to overreact on news in a short period of time.

Originality/value – While sentiment analysis of Twitter posts as a predictor of the Bitcoin price has been conducted in the past, this research does not have any analog because psycho-semantic dictionaries have not been applied earlier in the Bitcoin research.

Keywords Bitcoin, Market efficiency, Cryptocurrencies, Lexicon-based sentiment analysis

Paper type Research paper

1. Introduction

In public and academia, there is an ongoing debate about what constitutes the value of Bitcoin and whether it is anything but a hype or a bubble. In this context, it is needless to say that ample research opportunities exist to better understand the drivers of the Bitcoin price. Past research focused on fundamental and technical analysis to predict price variations and reveal causal relationships. Furthermore, there exists academic literature investigating the more social aspects of Bitcoin and its price formation. This research will continue along the same line and will test whether the value of Bitcoin is largely determined by sentiment, in contrast to more fundamental drivers.

1.1 Sentiment-based price predictions

Tetlock (2007) was the first to measure the interactions between (mass) media reporting and financial market movements. The goal of this study is to systematically analyze whether sentiment can be used as one of the predictors of securities' price movements. To do so, a lexicon-based sentiment analysis is applied. Using a tool called the General Inquirer[1], the polarity of the Abreast of the Market column, published daily in *The Wall Street Journal*, is assessed. This column gives a detailed discussion on yesterday's stock market happenings and, from time to time, tries to predict what the market will do in the near future. First, the principal component analysis is used to construct a simple measure of media pessimism.



Then, vector autoregressions (VARs) are used to estimate the intertemporal link between media pessimism and the stock market. Tetlock found that high levels of media pessimism induce a downward pressure on the market prices, followed by a reversion to the fundamentals. The opposite turns out to be true as well: low market returns precede high media pessimism. Also, extreme media pessimism is followed by higher trading volumes. He concluded that media content mainly serves as a proxy for investor sentiment and not for new fundamental information or market volatility. He proposed a strategy that would yield nontrivial excess returns, though it remains unclear if that strategy would still be profitable after accounting for commissions, bid-ask spreads, taxes, etc.

Loughran and McDonald (2011) were the first to describe the potential benefits of using a domain-specific dictionary. Previous research relied on counting negative words for assessing the tone of financial writings. For that purpose, lexicons that were originally developed for other disciplines are often applied. Loughran and McDonald were against this practice and advised that the majority of negative words in the Harvard-IV-4 TagNeg word list should not be considered negative in a financial context, for example, cancer in cancer research. Thus, they proposed a new, manually constructed, finance-oriented dictionary[2]. When testing this context-specific lexicon on a sample of 10-Ks, they found significant relationships between the tone of the report and the file date returns, trading volume and subsequent return volatility.

Engelberg and Parsons (2011) found evidences of the causal effect of media sentiment on trading by identifying that a link between media sentiment and trading activity is broken on the days when newspapers are not delivered because of the natural disasters.

1.2 Past research to explain formation of Bitcoin price

As Bitcoin is a pioneer of the new asset class, namely, cryptocurrencies, there are a number of studies trying to identify factors determining its price formation. Alstynes (2014) identified the following four areas, which might explain the Bitcoin price formation: technical value determined by the consensus mechanism to ensure permanent public record, very low transaction fees, zero transaction fraud risk and value given by end users.

Yermack (2013) assessed whether Bitcoin's price behaves in the same manner as other currencies' and identified a very high volatility, which might be an obstacle for adoption of Bitcoin as a currency, and also found that it possesses a very low correlation with other currencies.

Hayes (2016) identified that Bitcoin's and other cryptocurrencies' prices are determined by the relative differences in the cost of production (mining) on the margin. It has also been shown that the value of cryptocurrency also depends on the cryptographic algorithms used.

The methods used in these studies have shown valid ways to address the Bitcoin value, which may not tell the whole story. In this paper, I seek to examine whether user sentiment in any way fills in the unexplained variation in price.

1.3 Attempts to use sentiment analysis to predict Bitcoin price movements

In addition, there are a number of quantitative research works that use investors' behavior to predict the Bitcoin price by looking at the search volumes, the social media shares or mentions of Bitcoin in media. Such studies include Kristoufek (2013), who was one of the first researchers to examine the relationship between search queries, as a proxy for investor interest, and the Bitcoin price. His research showed that there exist strong correlations between the Mt. Gox price data, the daily Wikipedia and the weekly Google search volumes. Garcia et al. (2014) tried to answer how social interactions play a role in the creation of

bubbles. Similar to Kristoufek's study, they used price data from Mt. Gox, Google and Wikipedia search volumes. They modeled information sharing, so-called online word-of-mouth, using the fraction of tweets related to Bitcoin and validated this measure using the Facebook reshares. The number of Bitcoin client downloads and unique users' downloads, proxied using heuristics on the blockchain, was used to estimate the user base. They found that there exist two feedback loops, namely, a social cycle and a user adoption cycle, where search volume reinforces the price via word-of-mouth and user adoption, respectively.

Only a small number of studies have considered a sentiment of publicly available textual information as an indicator for the Bitcoin price movements. [Garcia and Schweitzer \(2015\)](#) revealed that hikes in emotional valence precede increases in opinion polarization and exchange volume. In turn, an increase in these factors precedes higher Bitcoin prices. Emotional valence and opinion polarization of tweets were measured using psycholinguistic lexicon-based methods. A VAR model was applied and the findings were used to design an algorithmic trading strategy. This strategy was benchmarked against both simple and random strategies. Even when taking into account the transaction costs, they generate high profits at acceptable risks. This leads to a Sharpe ratio of 1.823. In a research conducted by [Kaminski and Gloor \(2014\)](#), Twitter data of 104 days were mined in search of positive, negative and uncertain Bitcoin-related tweets. To do so, a very small lexicon (15 words) was used. They showed that negative tweets and tweets that express uncertainty correlate moderately positive with the Bitcoin trading volume and negatively with the Bitcoin price. Introducing lag shows that high trading volumes precede a rise in such tweets. Although a Granger causality analysis could not confirm any causal effect, they concluded that Twitter sentiment only mirrors and not predicts the market. [Polasik et al. \(2015\)](#) went further and performed a lexicon-based sentiment analysis on English articles mentioning the word Bitcoin in the Nexis database. They also included a number of such articles and the Google search volume as proxies for the popularity of this currency. Using regressions they found that the tone of press articles and the popularity of Bitcoin are the strong factors. According to their analysis, increases in transaction volume lead to a higher price, whereas the global economic climate does not seem to be an important driver. The previous studies regarding sentiment analysis excluded Bitcoin-specific news outlets and used very small lexicons.

2. Data and methodology

2.1 Data collection

Various Bitcoin-related news portals were used to perform the sentiment analysis experiments. All input came from a database containing 15,850 articles. [Appendix 1](#) gives an overview of the sources contributing to this database. Bitcoin expert media articles were collected by scraping these sources from their websites and parsing was done using Beautiful Soup[3]. For data collected using scrapers, the time frame begins with the launch of the news website and ends around the beginning of February 2016. News stories on websites dedicated to Bitcoin, even if not read by all Bitcoin traders, cover most publicly available Bitcoin-related information and hence, is a good proxy for a media sentiment surrounding the Bitcoin. As a consequence, for these sources no article selection has taken place except for the filtering of articles with missing elements. This article set was used to find the most frequent FinTech-related words. These shortlisted words[4] were then used to construct the final article set. All incomplete, for example a missing date or headline, articles were removed. The articles that turned out to be irrelevant after all, e.g. the article containing the words virtual and currency but not consecutively, were removed as well. As a last step, duplicate articles were filtered out. Also, in case where two articles turned out to be very similar, e.g. a corrected typo in the body, one of those was deleted from the set. Price

data were obtained from the Bitstamp exchange[5]. To make sure that the articles and the price were linked correctly, all dates and times were converted to GMT + 0 (Figure 1).

All posts and articles were preprocessed in a similar fashion. First, they were broken down into sentences and then the sentences into words. Using a part-of-speech tagger, provided by the Natural Language Toolkit [6], the words were tagged. Then, the WordNet Lemmatizer was used to get the correct lemma taking into account its grammatical category.

2.2 Choosing a dictionary

Lexicon-based techniques determine the orientation of a document by evaluating the words written against a sentiment or subjectivity lexicon. The underlying assumption is that the collective polarity of a piece of text is the sum of the sentiment scores of the individual words (Kaushik and Mishra, 2014). For example, a piece of text with a high number of positive words relative to the number of negative words will be considered optimistic. Both Loughran and McDonald (2011) and Siering (2012) showed that the performance of such an exercise drastically improves when one uses a context-specific dictionary. General lists are proven to more often mistakenly identify the connotation or context of a word (Grayson *et al.*, 2014; Soroka *et al.*, 2015). Therefore, in addition to the Harvard Psychosocial Dictionary, I used a finance-focused lexicon created by Loughran and McDonald (2011). After comparing these dictionaries with regard to the size and overlap, I found that only 508 words out of 3,626 included in the Harvard dictionary and 2,709 included in Loughran–McDonald dictionary overlap. Lexicons performance is measured in terms of correct predictions made by the sentiment analyzer. To decide whether a sentiment prediction is correct, price data are used. Thus, the best-performing dictionary is the one that yields the highest returns.

2.3 Sentiment analysis

Sentiment analysis is a subfield of natural language processing. Sentiment analysis attempts to extract subjective information from not only the textual documents, such as news articles and product reviews, but also voice and video. Sentiment classification is a typical job, which includes identifying the polarity, e.g. positive or negative, of a document. A major advantage when applying sentiment analysis to the task of predicting or forecasting the rate of a financial asset is that one can measure and judge the impact of a variety of events without the need of specifying those events (Tetlock *et al.*, 2007). One way to implement the sentiment analysis is a lexicon-based approach. Such a technique uses a subjectivity lexicon: a list of words and term weights that expresses the sentiment polarity of the words, as discussed in Subsection 2.2. Typically, the outcome of a sentiment orientation exercise will either be a classification, e.g. positive, negative or neutral, or a

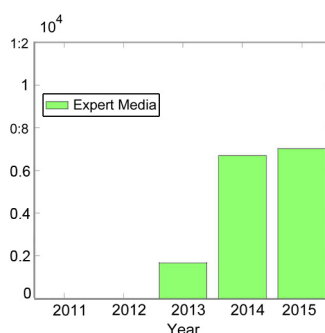


Figure 1.
Number of articles
per year, articles
before 2011 are
omitted

sentiment value within a certain range, e.g. from -1 (very negative) to 1 (very positive) (Serrano-Guerrero *et al.*, 2015). Document-level sentiment classification will output one classification or a single score for the entire document. This might be wrong as, for example, a news article can express multiple views. Segment-level opinion analysis tries to address these concerns (Cambria *et al.*, 2013). This paper used lexicon-based sentiment analysis on the document level with generic and field-specific dictionaries; sentiment scores are within the $[-1,1]$ range. Because existing tools generally function as a black box, their parametrization and/or the use of a custom dictionary proves to be difficult. Therefore, I built my own sentiment analyzer using the best practices found both in literature and on the Web.

Automated sentiment detection boils down to three steps: feature extraction, feature scoring and score aggregation. Features that are commonly used in documents are term presence, term frequency, the position of a word and negations (Medhat *et al.*, 2014). Bag of words is the simplest model, in which a document is treated as a multiset of its words. In this paper, I used a slightly more sophisticated approach. A document that was fed to our sentiment analyzer was considered as a set of sentences. All words of each sentence were compared to three lists: a list of negators, a list of intensifiers and a sentiment lexicon. The list of intensifiers[7] and negators[8] were copied from the study conducted by Loughran and McDonald (2011) and the study conducted by Brooke (2009), respectively. While the negators simply flip the polarity, intensifiers add or subtract some weight. These weights were also given by Brooke and are the work of several researchers. Whenever a word shows up in the sentiment dictionary, its term weight is added to the tally keeping a track of the positive (negative) score in case of a positive (negative) word. Although the local context of the word is considered, if a contextual valence shifter precedes the dictionary word, its term weight can be amplified or downtoned, inverted or both. To do so, I used a four-word window, in line with the existing research (Loughran and McDonald, 2011). Consider the following example sentence: Blockchain technology is very popular nowadays. Let us assume the word popular shows up in our dictionary with a moderately positive weight of 0.9 . The words in our four-word window are (technology, is, very, popular), of which the word very shows up on the list of intensifiers. According to this list, the term weight should be amplified with 25 per cent. As a result, I will add 0.9×1.25 to the positive tally and move on to the next sentence. An illustration of a more complex example: I really do not like Bitcoin, it is the stupidest idea ever! Let us assume that both like and stupid are dictionary words with respective weights of 1.2 and -1.4 . When processing the word like, the window is as follows: (really, do, not, like), which also contains really and not. The former will amplify the term weight by 15 per cent, and the latter will flip the polarity. As a consequence, $1.2 \times 1.15 \times -1$ will be added to the negative tally. Then the window slides further up until it hits the next dictionary word, the current state now is (it, is, the, stupidest). In this case, there are no contextual valence shifters: -1.4 will be added to the negative tally. Now the next sentence will be analyzed. Note that certain intensifiers will both flip and amplify the polarity of a dictionary word at the same time, e.g. the least. As I believed the headline of an article to be of greater importance than the body, I processed both parts separately. To come to the final number of positive and negative terms, the headline tallies were multiplied with an important factor and were added to the body tallies. I chose an important factor of five, meaning that I believed a positive or negative word in an article headline to have five times as much impact on the overall sentiment of the article than a positive or negative word in the body of the article would have. This was a rather arbitrary choice. In short, the sentiment analyzer takes as input a headline and body and returns two tallies containing the weighted number of positive and negative references.

Now these two scores need to be combined to get to a single sentiment measure. This is called a combining function (Jurek *et al.*, 2015). Many such functions exist, one combining function that is often used for document-level sentiment analysis is:

$$\text{sentiment score} = \frac{P - N}{P + N} \quad (1)$$

with P and N the positive and negative tally, respectively (Siering, 2012; Tetlock *et al.*, 2007). As was mentioned in Subsection 2.2, using the term presence rather than frequency might give a more nuanced view on the sentiment expressed in a piece of text. This was also shown by Pang *et al.* (2002). As a result of these findings, I opted to use the term presence, meaning that every dictionary word that occurs in a document will add to a tally only once. However, I had to take into account the local context, e.g. it might be inverted or amplified differently in different parts of the text. To cope with this, I decided to consider only the first occurrence of a dictionary word, together with its local context, and ignore all further occurrences.

3. Experiments and results

To answer the research questions raised in this paper, I will take the interday perspective assuming that Bitcoin price does not incorporate new information immediately after its publication (Figure 2). This assumption is mainly based on the fact that in currency deposits to Bitcoin, it takes a day for the exchanges to be credited to the trader's account as funds are usually deposited via bank wires. Interday boils down to swing trading. Also I will neglect position trading, as I believe it can only take a few days for the market to act upon new information. The challenges posed by interday trading will be discussed separately. The goal of this section is to test whether an intelligent investor could have made abnormal results by leveraging media sentiment.

3.1 Interday trading

The first step is to see how and whether the market responds to a change in media sentiment. One could assume the price to simply follow the sentiment movements. However, this would ignore all kinds of market reactions a swing trader could exploit. So, it is also important to test for more complex temporal patterns, such as delayed reactions, under- and overreactions and their resulting corrections. To visualize these movements, I create reaction patterns. These show the average relative price movement, i.e. the return over that period, after a few noteworthy news stories. To make these graphs, I first need to define what can be considered noteworthy. To do so, I introduce the following three variables: look-up period, lag period and hold period. The look-up period is the number of days over which the article scores are aggregated. The real world equivalent is how often one reads the news. This defines the considered time frame to decide whether something noteworthy is happening within the FinTech world or not. If so, the price movements following this period will be measured. However, as was mentioned before, one cannot expect the market to move



Figure 2.
Timeline
representing the three
variables: look-up, lag
and hold

immediately. The lag period defines how many days it takes for the price to start adjusting according to the change in media sentiment. This lag can occur because of the investors reading the news with delay, factors that inhibit instant trading, personal or strategic choices, etc. The hold period is the duration one should hold on to his/her position to maximize the captured price change. In other words, the time it would take before the sentiment information is fully incorporated or before a reaction inverts the price movement.

The usage of look-up period eliminates the risk of news duplicates (the cases when the same news is republished in a different outlet) as the sentiment used for investment decision is based on a media sentiment given a specific period of time, and not on each article itself. Using our three-variable model, finding out how the market responds to a change in media sentiment is equivalent to searching for a variable combination that consistently predicts strong price movements. As many possible combinations exist, I plot the average single-day return for a sample mix of lag and look-up combinations.

To decide when a market movement can be expected, I compute the aggregated sentiment over the look-up period. First, all articles published during the look-up period need to be retrieved. Second, the sentiment of these articles is computed using our sentiment analyzer. Finally, these individual scores are aggregated by taking a simple sum. However, not all articles or posts are taken into account. Neutral texts tend to fool many sentiment analyzers (Serrano-Guerrero *et al.*, 2015). In such texts, some sentiment words will still be used according to the author's writing style. In case one uses a sentiment metric that works with a ratio of the positive and negative word occurrences, e.g. equation (1), he/she might mistakenly classify neutral texts as (very) positive or negative. To prevent this, it is important to set a threshold on the number of sentiment words that minimally have to occur as well as a minimal number of words in the article. The choice for a minimum of fifty words and at least five sentiment words was inspired by a previous research (Tetlock *et al.*, 2007). This same source also enforces at least three unique sentiment words. Given our choice to use the term presence, all counted sentiment words are unique. Because of our variable term weight, the threshold is more precisely defined as the sum of the absolute values of the modified term weights that has to be greater than or equal to five. The rationale for these constraints is that it is probably better to exclude many articles than to make trades based on articles that express no sentiment at all.

A perfect sentiment analyzer would return a positive or negative score for all positive or negative articles or posts, respectively. Though, this is almost never the case. A dictionary might have more positive or negative words. Also, authors could on average write with a more positive or negative vocabulary. As a consequence, a sentiment value of zero might not necessarily be the cut-off between positive and negative words. Plus it could be that there is some long-term sentiment trend, e.g. the media in general reports negatively about Bitcoin. A sustained negative sentiment is unlikely to induce any price movement, and only relative changes will do so. I will measure this baseline by keeping a track of a simple moving average of the sentiment score, computed since the beginning of 2014. This average is updated daily and deducted from the sentiment score of that day. As the last step, I have to define the deviation from the baseline I thought necessary for the media to induce market movements. The rule used in our model is that the sentiment over the look-up period has to be more than one below or above the sentiment baseline. The rationale for choosing one is that this is equivalent to requiring more than one extremely positive or negative article or posts on top of the baseline. While a single article or post could simply be a one-man rage, starting from more than one I might assume that sentiment is shifting one way or the other.

Simulations were run for each day of the year to ensure that I capture all kinds of market conditions[9]. The average relative price movement is then computed. For return

calculations, the Bitstamp closing price is used. In case no price data are available or Bitstamp service is suspended, I skip trading those days.

3.1.1 Prediction scenarios. As finding a sentiment-induced price pattern is not the same as having a workable strategy, I will now try to put these insights into practice. The goal is to see whether such a strategy could yield high enough returns to compensate for the risk you take [10]. If so, then expert media would clearly predict and may even influence the market. In case this strategy yields abnormal results based on public information, I will also be able to conclude that the Bitcoin market is not semi-strongly efficient. Prediction scenarios are a form of backtesting where one uses a system to predict historic price movements. These scenarios serve to measure the accuracy and profitability as a way to validate the system (Garcia and Schweitzer, 2015). An intelligent investor who hypothesized or noticed that sentiment expressed in expert media is followed by a recurring price pattern might have been able to make good returns. To estimate how well off such an investor would have been, I simulate a more complex sentiment trading strategy that tries to capture all the price swings. I then compute and compare the returns and the Sharpe ratios and perform some statistical tests.

The proposed trading strategy goes as follows. An investor assesses the news every day, and at the end of the day he makes the decision to act or not act. In case of positive signals, he will go long the first day, whereas negative signals will induce a short position. After capturing the initial reaction, the investor will do the exact opposite for the next two days in an attempt to catch the corrective price movement. Finally, he will revert his position again hoping that increased market liquidity steers the price to a more long-term trend. To assess the performance of this strategy under various market conditions, simulations with different durations and starting points were run. Care was taken to ensure the investor uses nothing more than data collected within the specified time window, historic data and the knowledge of the earlier mentioned pattern. This strategy specifies the Bitcoin position to be held for five days after the sentiment signal. Given that we make such a sentiment evaluation on a daily basis, it is possible to hold multiple positions at a time. Because of these overlapping trades, an investor will cancel out some of his positions. For example, assume the outcome of the algorithm is such that you have to keep two long positions and one short position, and then you will effectively have only one long position. As a result, you will capture only one third of the return of that day. More formally:

$$\frac{L - S}{L + S} R_{\text{daily}} \quad (2)$$

with L and S representing the number of long and short positions, respectively, and R_{daily} the daily return. By doing so, we make abstraction of the amount of money invested, which is often hard to quantify for market-making strategies. Other researchers (Shah and Zhang, 2014) coped with this issue by taking the average investment. However, one can argue that returns based on average investment exaggerate the performance of a portfolio.

Statistics and performance metrics are summarized in Tables I and II, respectively. Once again, Bitstamp data are used. To replicate these results, one needs to use the closing price, so in practice one would have to trade at the end of the day. Note that Bitstamp allows for buying or selling instantly. However, it does not allow for shorting. So, we have to assume that you can borrow Bitcoins from, for example, a long-term trader and then sell and re-buy on the Bitstamp exchange. The returns reported here are compounded. Both periodic and annualized returns are displayed to ease comparison between time windows with different length. Longer periods perform in the double-digit range. Shorter periods appear to be more

Table I.
Performance of the
interday sentiment
strategy using
Harvard
Psychosocial
Dictionary

Start	End	Days	Articles	Start price	End price	Annual return	Sharpe	Meaning
01/01/14	31/12/15	730	11,488	754.22	430.89	0.45	1.14	Good
01/01/14	31/12/14	365	5,689	754.22	321.00	0.70	1.49	Good
01/01/15	31/12/15	365	5,799	313.81	430.89	0.10	0.45	Poor
01/01/14	30/06/14	181	3,192	754.22	641.11	0.60	1.14	Good
01/07/14	31/12/14	184	2,497	638.25	321.00	0.99	2.85	Very good
01/01/15	30/06/15	181	3,010	313.81	262.50	−0.26	−0.58	Poor
01/07/15	31/12/15	184	2,789	257.62	430.89	0.87	1.95	Good
01/01/14	31/03/14	90	1,449	754.22	454.83	−0.60	−1.27	Poor
01/04/14	30/06/14	91	1,743	478.98	641.11	3.38	3.08	Excellent
01/07/14	30/09/14	92	1,363	638.25	391.00	0.51	1.94	Good
01/10/14	31/12/14	92	1,134	384.86	321.00	0.88	2.61	Very good
01/01/15	31/03/15	90	1,458	313.81	244.24	−0.67	−2.00	Poor
01/04/15	30/06/15	91	1,552	246.69	262.50	0.23	1.03	Good
01/07/15	30/09/15	92	1,573	257.62	236.20	0.03	0.22	Poor
01/10/15	31/12/15	92	1,216	237.15	430.89	1.55	2.46	Very good
01/01/16	31/03/16	91	1,248	433.82	415.35	−0.36	−1.37	Poor
06/05/14	24/06/14	50	943	428.01	572.62	0.20	0.66	Poor

Notes: All returns are compounded. Annual return shows the annualized return. For example, for the first quarter of 2016, the strategy would have yielded 36% annual loss. The Sharpe ratio of −1.37 indicates that we are taking excessive risks

Table II.
Performance of the
interday sentiment
strategy using
Loughran–McDonald
finance-specific
dictionary

Start	End	Days	Articles	Start price	End price	Annual return	Sharpe	Meaning
01/01/14	31/12/15	730	11,488	754.22	430.89	0.82	1.84	Good
01/01/14	31/12/14	365	5,689	754.22	321.00	1.28	2.23	Very good
01/01/15	31/12/15	365	5,799	313.81	430.89	0.32	1.03	Good
01/01/14	30/06/14	181	3,192	754.22	641.11	0.77	1.47	Good
01/07/14	31/12/14	184	2,497	638.25	321.00	2.06	3.31	Excellent
01/01/15	30/06/15	181	3,010	313.81	262.50	−0.15	−0.28	Poor
01/07/15	31/12/15	184	2,789	257.62	430.89	1.10	2.62	Very good
01/01/14	31/03/14	90	1,449	754.22	454.83	1.20	1.99	Good
01/04/14	30/06/14	91	1,743	478.98	641.11	1.94	2.39	Very good
01/07/14	30/09/14	92	1,363	638.25	391.00	−0.40	−3.14	Poor
01/10/14	31/12/14	92	1,134	384.86	321.00	9.45	5.46	Excellent
01/01/15	31/03/15	90	1,458	313.81	244.24	−0.07	0.07	Poor
01/04/15	30/06/15	91	1,552	246.69	262.50	−0.17	−0.61	Poor
01/07/15	30/09/15	92	1,573	257.62	236.20	1.85	4.16	Excellent
01/10/15	31/12/15	92	1,216	237.15	430.89	0.59	1.55	Good
01/01/16	31/03/16	91	1,248	433.82	415.35	1.71	2.51	Very good
06/05/14	24/06/14	50	943	428.01	572.62	0.01	0.16	Poor

Notes: All returns are compounded. Annual return shows the annualized return. For example, for the first trimester of 2016, the strategy would have yielded more than 171% annual return. The Sharpe ratio of 2.51 indicates that we do not take excessive risks

volatile resulting in two scenarios with negative results. The lowest return of strategy using the Harvard Psychosocial Dictionary, e.g. an annual loss of 60 per cent, coincides with the collapse of the Bitcoin price. It seems that if the market is crashing down, the algorithm cannot save you:

$$\begin{aligned} \text{Sharpe ratio} &= \text{SR}(r \sim_e) = \frac{E[r \sim_e]}{\sqrt{\text{Var}[r \sim_e]}} \sqrt{365} \text{ with } r \sim_e \\ &= (r_{e1}; \dots; r_{en}) \text{ and} \end{aligned}$$

$$\text{PV}[r \sim_e] r_{ei} \text{ the excess return on the } i^{\text{th}} \text{ day} \quad (3)$$

However, comparing returns without risk is meaningless, especially when a volatile asset such as Bitcoin is combined with a complex trading strategy. To incorporate a risk, Sharpe ratios (Sharpe, 1994) are computed. The average daily excess return is divided by its standard deviation resulting in a daily Sharpe, which I then annualize for the sake of comparison[11]. In accordance with other researchers, the current risk-free return is assumed to be zero (Garcia and Schweitzer, 2015)[12]. It can be argued that asset-specific risk-free rate should be used. In Bitcoin case, the Bitcoin-specific risk-free rate might be deducted from Bitcoin lending rates at various exchanges offering margin lending facilities; however, these rates mainly represent the risk premium for holding Bitcoins at the centralized facility, which is usually neither regulated nor insured. Also, a large fraction[13] of Bitcoins is held as investment or savings, hence a rational holder would prefer to invest at a risk-free rate given it is higher than zero. Hence, because of high uncertainty of Bitcoin price, the demand for Bitcoin-denominated loans should also be very low. As a result, it can be assumed that the Bitcoin-specific risk-free rate is zero.

The last columns of Tables I and II apply a rule of thumb to give meaning to these Sharpe indexes[14]. Further details are given in Tables III and IV, which compare the returns and Sharpe ratios for two major stock market indexes to the ones obtained using

Start	End	Harvard	Annualized compounded return		Harvard	Sharpe ratio	
			DJIA	S&P500		DJIA	S&P500
01/01/14	31/12/15	0.45	0.03 (0.10)	0.06 (0.13)	1.14	0.28 (0.17)	0.48 (0.25)
01/01/14	31/12/14	0.70	0.08 (0.13)	0.12 (0.16)	1.49	0.80 (0.34)	1.09 (0.45)
01/01/15	31/12/15	0.10	-0.02 (0.36)	-0.01 (0.37)	0.45	0.04 (0.36)	0.08 (0.38)
01/01/14	30/06/14	0.60	0.05 (0.25)	0.15 (0.31)	1.14	0.51 (0.39)	1.37 (0.61)
01/07/14	31/12/14	0.99	0.10 (0.07)	0.09 (0.07)	2.85	0.44 (0.07)	0.55 (0.08)
01/01/15	30/06/15	-0.26	-0.02 (0.64)	0.00 (0.65)	-0.58	0.04 (0.65)	0.08 (0.66)
01/07/15	31/12/15	0.87	-0.04 (0.14)	-0.03 (0.14)	1.95	0.12 (0.12)	0.08 (0.12)
01/01/14	31/03/14	-0.60	0.00 (0.73)	0.09 (0.76)	-1.27	0.09 (0.70)	0.83 (0.81)
01/04/14	30/06/14	3.38	0.07 (0.08)	0.17 (0.10)	3.08	0.72 (0.21)	1.52 (0.32)
01/07/14	30/09/14	0.51	0.02 (0.31)	0.00 (0.29)	1.94	0.76 (0.32)	0.64 (0.30)
01/10/14	31/12/14	0.88	0.26 (0.15)	0.25 (0.16)	2.61	-0.22 (0.09)	-0.09 (0.11)
01/01/15	/03/15	-0.67	-0.01 (0.82)	0.02 (0.82)	-2.00	0.04 (0.80)	0.07 (0.80)
01/04/15	30/06/15	0.23	-0.02 (0.37)	0.01 (0.37)	1.03	0.05 (0.33)	0.06 (0.34)
01/07/15	30/09/15	0.03	-0.29 (0.49)	-0.27 (0.48)	0.22	0.12 (0.49)	0.08 (0.48)
01/10/15	31/12/15	1.55	0.31 (0.21)	0.27 (0.20)	2.46	1.11 (0.31)	0.93 (0.28)
01/01/16	31/03/16	-0.36	0.13 (0.74)	0.10 (0.72)	-1.37	0.43 (0.80)	0.26 (0.78)
06/05/14	24/06/14	0.20	0.20 (0.42)	0.37 (0.44)	0.66	0.23 (0.46)	0.49 (0.49)

Table III.
Benchmark of the
interday sentiment
strategy using
Harvard's dictionary
versus market
indexes

Notes: The p -values of H_0 $\$R_{\text{sent}} R_{\text{index}}$ and H_0 $\$SR_{\text{sent}} SR_{\text{index}}$ are added between parentheses, R expresses the average daily return and SR the daily Sharpe. These p -values indicate the likeliness that an index yields higher returns or Sharpe ratios than the sentiment strategy. It can be seen that neither in terms of return nor Sharpe ratio the sentiment strategy using Harvard's dictionary outperformed the market

Table IV.
Benchmark of the
interday sentiment
strategy using
Loughran–
McDonald’s
dictionary versus
market indexes

Start	End	Loughran– McDonald	Annualized compounded return		Loughran– McDonald	Sharpe ratio	
			DJIA	S&P500		DJIA	S&P500
01/01/14	31/12/15	0.82	0.03 (0.02)	0.06 (0.03)	1.84	0.28 (0.03)	0.48 (0.06)
01/01/14	31/12/14	1.28	0.08 (0.04)	0.12 (0.05)	2.23	0.80 (0.16)	1.09 (0.23)
01/01/15	31/12/15	0.32	−0.02 (0.21)	−0.01 (0.21)	1.03	0.04 (0.19)	0.08 (0.20)
01/01/14	30/06/14	0.77	0.05 (0.19)	0.15 (0.25)	1.47	0.51 (0.32)	1.37 (0.54)
01/07/14	31/12/14	2.06	0.10 (0.03)	0.09 (0.03)	3.31	0.44 (0.05)	0.55 (0.06)
01/01/15	30/06/15	−0.15	−0.02 (0.57)	0.00 (0.58)	−0.28	0.04 (0.58)	0.08 (0.59)
01/07/15	31/12/15	1.10	−0.04 (0.09)	−0.03 (0.09)	2.62	0.12 (0.05)	0.08 (0.05)
01/01/14	31/03/14	1.20	0.00 (0.18)	0.09 (0.22)	1.99	0.09 (0.23)	0.83 (0.35)
01/04/14	30/06/14	1.94	0.07 (0.14)	0.17 (0.17)	2.39	0.72 (0.28)	1.52 (0.42)
01/07/14	30/09/14	−0.40	0.02 (0.91)	0.00 (0.90)	−3.14	0.76 (0.96)	0.64 (0.95)
01/10/14	31/12/14	9.45	0.26 (0.01)	0.25 (0.01)	5.46	−0.22 (0.01)	−0.09 (0.01)
01/01/15	31/03/15	−0.07	−0.01 (0.49)	0.02 (0.49)	0.07	0.04 (0.50)	0.07 (0.50)
01/04/15	30/06/15	−0.17	−0.02 (0.59)	0.01 (0.59)	−0.61	0.05 (0.62)	0.06 (0.62)
01/07/15	30/09/15	1.85	−0.29 (0.08)	−0.27 (0.08)	4.16	0.12 (0.04)	0.08 (0.04)
01/10/15	31/12/15	0.59	0.31 (0.37)	0.27 (0.36)	1.55	1.11 (0.46)	0.93 (0.42)
01/01/16	31/03/16	1.71	0.13 (0.17)	0.10 (0.16)	2.51	0.43 (0.19)	0.26 (0.17)
06/05/14	24/06/14	0.01	0.20 (0.50)	0.37 (0.52)	0.16	0.23 (0.52)	0.49 (0.56)

Notes: The p -values of H_0 $\$R_{\text{sent}} R_{\text{index}}$ and H_0 $\$SR_{\text{sent}} SR_{\text{index}}$ are added between parentheses, R expresses the average daily return and SR the daily Sharpe. These p -values indicate the likeliness that an index yields higher returns or Sharpe ratios than the sentiment strategy. For example, during the second half of 2014, we can see that both in terms of return and Sharpe ratio, the sentiment strategy using Loughran–McDonald dictionary outperformed the market

sentiment strategies[15]. One might say that benchmarking Bitcoin price against a market index is not as insightful as with traditional financial assets. Two statistical tests have been added: Welch’s t -test for comparing the returns and a Z -test using the distribution derived from the study conducted by Jobson and Korkie (1981) for comparing the Sharpe ratios[16]. Both cases assume normally distributed returns, as is often done in practice. However, this ignores serial correlation, time-varying conditional volatilities and other non-iid[17] behavior (Opdyke, 2006). One might believe these p -values to be low, though one has to take into account that we are testing if the average daily return or Sharpe of our Bitcoin strategy is lower than those of an index. As the latter are more often than not non-negative, this implies testing whether your strategy significantly yields non-negative average daily returns. Thus, it is statistically improbable to have a period, for example 90 days, with a negative average daily return, which is quite an impossible feat. The rule of thumb and the stock index comparison both show that, also after incorporating the risk, the sentiment strategy obtains favorable results.

Tables V–VIII summarize the returns and Sharpe ratios of various simple, random and technical strategies applied to Bitcoin (Garcia and Schweitzer, 2015) and compare those to the strategy proposed in this paper. Buy and hold is the simplest: one buys a Bitcoin at the beginning of the time window and sells it at the end. In the random strategy, one flips a coin to decide whether he will go long or short that day. Executing the momentum strategy means that you believe today’s price movement to be in line with the movement of the day before. The up and down persistency (UDP) strategy suggests to do the exact opposite: always go against yesterday’s price movement. Momentum is the strategy that comes closest to the sentiment strategy, yielding higher returns in four of the seventeen periods.

Interday Bitcoin price movements

Start	End	Harvard	No short	Buy and hold	Random	Momentum	UDP
01/01/14	31/12/15	0.45	-0.33 (0.14)	-0.27 (0.19)	-0.08 (0.32)	0.17 (0.48)	-0.50 (0.07)
01/01/14	31/12/14	0.70	-0.57 (0.07)	-0.58 (0.08)	-0.63 (0.06)	1.30 (0.72)	-0.76 (0.02)
01/01/15	31/12/15	0.10	0.12 (0.59)	0.27 (0.66)	-0.42 (0.27)	-0.37 (0.31)	-0.03 (0.52)
01/01/14	30/06/14	0.60	-0.32 (0.33)	-0.41 (0.31)	-0.66 (0.19)	0.05 (0.46)	-0.58 (0.23)
01/07/14	31/12/14	0.99	-0.73 (0.01)	-0.75 (0.01)	-0.75 (0.01)	3.39 (0.85)	-0.83 (0.00)
01/01/15	30/06/15	-0.26	-0.60 (0.37)	-0.47 (0.47)	-0.59 (0.39)	-0.33 (0.54)	-0.18 (0.60)
01/07/15	31/12/15	0.87	1.75 (0.70)	1.67 (0.68)	-0.44 (0.14)	-0.40 (0.15)	0.16 (0.35)
01/01/14	31/03/14	-0.60	-0.84 (0.37)	-0.85 (0.38)	-0.87 (0.36)	1.66 (0.83)	-0.84 (0.38)
01/04/14	30/06/14	3.38	2.31 (0.49)	1.15 (0.40)	2.62 (0.50)	-0.46 (0.17)	-0.10 (0.25)
01/07/14	30/09/14	0.51	-0.85 (0.02)	-0.84 (0.02)	2.52 (0.82)	1.79 (0.75)	-0.71 (0.07)
01/10/14	31/12/14	0.88	-0.52 (0.17)	-0.54 (0.18)	0.08 (0.38)	5.65 (0.86)	-0.90 (0.02)
01/01/15	31/03/15	-0.67	-0.76 (0.51)	-0.70 (0.56)	-0.53 (0.63)	-0.45 (0.66)	-0.38 (0.68)
01/04/15	30/06/15	0.23	-0.27 (0.28)	-0.06 (0.39)	-0.18 (0.33)	0.28 (0.53)	-0.31 (0.26)
01/07/15	30/09/15	0.03	-0.31 (0.38)	-0.32 (0.38)	-0.42 (0.33)	-0.76 (0.10)	2.30 (0.88)
01/10/15	31/12/15	1.55	10.32 (0.86)	8.89 (0.83)	-0.96 (0.01)	0.74 (0.44)	-0.64 (0.13)
01/01/16	31/03/16	-0.36	-0.19 (0.61)	-0.09 (0.64)	-0.41 (0.51)	-0.31 (0.56)	0.07 (0.69)
06/05/14	24/06/14	0.20	8.35 (0.86)	9.16 (0.87)	1.40 (0.66)	-0.36 (0.40)	0.00 (0.49)

Notes: The p -values of H_0 $\$R_{\text{sent}} R_{\text{other}}$ are added between parentheses, R expresses the average daily return. These p -values indicate the likeliness that an alternative strategy yields higher returns than the sentiment strategy. All returns are compounded and annualized. For example, for the first trimester of 2016, the interday strategy using Loughran–McDonald dictionary yields an annualized 171%, whereas all other strategies, except for up and down persistency, generate negative returns

Table V.
Return benchmark of
the interday
sentiment strategy
using Harvard's
dictionary versus
simple, random and
technical strategies

Start	End	Harvard	No short	Buy and hold	Random	Momentum	UDP
01/01/14	31/12/15	1.14	-0.23 (0.04)	-0.07 (0.08)	0.25 (0.11)	0.58 (0.22)	-0.58 (0.03)
01/01/14	31/12/14	1.49	-0.84 (0.01)	-0.78 (0.03)	-0.94 (0.02)	1.48 (0.50)	-1.49 (0.01)
01/01/15	31/12/15	0.45	0.51 (0.52)	0.70 (0.58)	-0.44 (0.23)	-0.31 (0.24)	0.31 (0.45)
01/01/14	30/06/14	1.14	-0.03 (0.19)	-0.14 (0.21)	-0.76 (0.08)	0.51 (0.31)	-0.52 (0.18)
01/07/14	31/12/14	2.85	-2.15 (0.00)	-2.21 (0.01)	-2.18 (0.00)	2.93 (0.52)	-2.93 (0.00)
01/01/15	30/06/15	-0.58	-0.82 (0.44)	-0.42 (0.53)	-0.73 (0.46)	-0.13 (0.62)	0.13 (0.65)
01/07/15	31/12/15	1.95	2.12 (0.54)	1.94 (0.50)	-0.65 (0.05)	-0.54 (0.07)	0.54 (0.17)
01/01/14	31/03/14	-1.27	-1.73 (0.41)	-1.54 (0.46)	-1.69 (0.44)	1.52 (0.97)	-1.52 (0.46)
01/04/14	30/06/14	3.08	1.85 (0.22)	1.33 (0.17)	1.94 (0.28)	-0.32 (0.06)	0.30 (0.11)
01/07/14	30/09/14	1.94	-3.84 (0.01)	-3.70 (0.01)	2.94 (0.70)	2.43 (0.61)	-2.43 (0.07)
01/10/14	31/12/14	2.61	-0.93 (0.10)	-0.94 (0.12)	0.43 (0.22)	3.38 (0.62)	-3.38 (0.01)
01/01/15	31/03/15	-2.00	-0.91 (0.67)	-0.63 (0.70)	-0.20 (0.81)	-0.06 (0.83)	0.06 (0.78)
01/04/15	30/06/15	1.03	-0.72 (0.17)	0.01 (0.30)	-0.36 (0.22)	0.86 (0.46)	-0.86 (0.22)
01/07/15	30/09/15	0.22	-0.59 (0.38)	-0.54 (0.40)	-0.87 (0.31)	-2.71 (0.14)	2.71 (0.90)
01/10/15	31/12/15	2.46	4.23 (0.78)	3.71 (0.69)	-4.54 (0.00)	1.14 (0.28)	-1.14 (0.05)
01/01/16	31/03/16	-1.37	-0.17 (0.72)	0.11 (0.73)	-0.70 (0.64)	-0.41 (0.67)	0.41 (0.77)
06/05/14	24/06/14	0.66	3.89 (0.89)	4.04 (0.90)	1.70 (0.66)	-0.39 (0.32)	0.32 (0.46)

Notes: The p -values of H_0 $\$SR_{\text{sent}} SR_{\text{other}}$ are added between parentheses, SR expresses the daily Sharpe. These p -values indicate the likeliness that an alternative strategy generates higher Sharpe ratios than the sentiment strategy. For example, for the first trimester of 2016, the interday strategy using Loughran–McDonald dictionary has a Sharpe ratio of 2.51. The second highest Sharpe (0.41) for that period is generated by the up and down persistency strategy

Table VI.
Sharpe benchmark of
the interday
sentiment strategy
using Harvard's
dictionary versus
simple, random and
technical strategies

Table VII.
Return benchmark of
the interday
sentiment strategy
using Loughran–
McDonald’s
dictionary versus
simple, random and
technical strategies

Start	End	Loughran– McDonald	No short	Buy and hold	Random	Momentum	UDP
01/01/14	31/12/15	0.82	−0.33 (0.07)	−0.27 (0.11)	−0.08 (0.20)	0.17 (0.34)	−0.50 (0.03)
01/01/14	31/12/14	1.28	−0.57 (0.03)	−0.58 (0.04)	−0.63 (0.03)	1.30 (0.60)	−0.76 (0.01)
01/01/15	31/12/15	0.32	0.12 (0.50)	0.27 (0.58)	−0.42 (0.20)	−0.37 (0.24)	−0.03 (0.44)
01/01/14	30/06/14	0.77	−0.32 (0.31)	−0.41 (0.29)	−0.66 (0.17)	0.05 (0.44)	−0.58 (0.21)
01/07/14	31/12/14	2.06	−0.73 (0.01)	−0.75 (0.01)	−0.75 (0.01)	3.39 (0.69)	−0.83 (0.00)
01/01/15	30/06/15	−0.15	−0.60 (0.33)	−0.47 (0.42)	−0.59 (0.35)	−0.33 (0.50)	−0.18 (0.57)
01/07/15	31/12/15	1.10	1.75 (0.67)	1.67 (0.65)	−0.44 (0.11)	−0.40 (0.12)	0.16 (0.31)
01/01/14	31/03/14	1.20	−0.84 (0.11)	−0.85 (0.13)	−0.87 (0.12)	1.66 (0.60)	−0.84 (0.14)
01/04/14	30/06/14	1.94	2.31 (0.57)	1.15 (0.48)	2.62 (0.59)	−0.46 (0.23)	−0.10 (0.31)
01/07/14	30/09/14	−0.40	−0.85 (0.10)	−0.84 (0.11)	2.52 (0.97)	1.79 (0.95)	−0.71 (0.26)
01/10/14	31/12/14	9.45	−0.52 (0.02)	−0.54 (0.02)	0.08 (0.08)	5.65 (0.40)	−0.90 (0.00)
01/01/15	31/03/15	−0.07	−0.76 (0.33)	−0.70 (0.38)	−0.53 (0.46)	−0.45 (0.48)	−0.38 (0.51)
01/04/15	30/06/15	−0.17	−0.27 (0.45)	−0.06 (0.57)	−0.18 (0.51)	0.28 (0.70)	−0.31 (0.43)
01/07/15	30/09/15	1.85	−0.31 (0.10)	−0.32 (0.11)	−0.42 (0.09)	−0.76 (0.02)	2.30 (0.58)
01/10/15	31/12/15	0.59	10.32 (0.93)	8.89 (0.91)	−0.96 (0.01)	0.74 (0.57)	−0.64 (0.20)
01/01/16	31/03/16	1.71	−0.19 (0.19)	−0.09 (0.23)	−0.41 (0.15)	−0.31 (0.18)	0.07 (0.27)
06/05/14	24/06/14	0.01	8.35 (0.90)	9.16 (0.91)	1.40 (0.71)	−0.36 (0.44)	0.00 (0.54)

Notes: The p -values of H_0 $\$R_{\text{sent}} R_{\text{other}}$ are added between parentheses, R expresses the average daily return. These p -values indicate the likeliness that an alternative strategy yields higher returns than the sentiment strategy. All returns are compounded and annualized. For example, for the first trimester of 2016, the interday strategy using Loughran–McDonald dictionary yields an annualized 171 %, whereas all other strategies, except for up and down persistency, generate negative returns

Table VIII.
Sharpe benchmark of
the interday
sentiment strategy
using Loughran–
McDonald’s
dictionary versus
simple, random and
technical strategies

Start	End	Loughran–McDonald	No short	Buy and hold	Random	Momentum	UDP
01/01/14	31/12/15	1.84	−0.23 (0.01)	−0.07 (0.02)	0.25 (0.02)	0.58 (0.04)	−0.58 (0.00)
01/01/14	31/12/14	2.23	−0.84 (0.00)	−0.78 (0.01)	−0.94 (0.01)	1.48 (0.22)	−1.49 (0.00)
01/01/15	31/12/15	1.03	0.51 (0.34)	0.70 (0.40)	−0.44 (0.12)	−0.31 (0.11)	0.31 (0.28)
01/01/14	30/06/14	1.47	−0.03 (0.14)	−0.14 (0.16)	−0.76 (0.06)	0.51 (0.23)	−0.52 (0.15)
01/07/14	31/12/14	3.31	−2.15 (0.00)	−2.21 (0.00)	−2.18 (0.00)	2.93 (0.40)	−2.93 (0.00)
01/01/15	30/06/15	−0.28	−0.82 (0.38)	−0.42 (0.47)	−0.73 (0.40)	−0.13 (0.54)	0.13 (0.58)
01/07/15	31/12/15	2.62	2.12 (0.39)	1.94 (0.37)	−0.65 (0.02)	−0.54 (0.04)	0.54 (0.09)
01/01/14	31/03/14	1.99	−1.73 (0.05)	−1.54 (0.11)	−1.69 (0.10)	1.52 (0.35)	−1.52 (0.13)
01/04/14	30/06/14	2.39	1.85 (0.37)	1.33 (0.28)	1.94 (0.41)	−0.32 (0.11)	0.30 (0.18)
01/07/14	30/09/14	−3.14	−3.84 (0.39)	−3.70 (0.42)	2.94 (1.00)	2.43 (1.00)	−2.43 (0.59)
01/10/14	31/12/14	5.46	−0.93 (0.00)	−0.94 (0.00)	0.43 (0.01)	3.38 (0.16)	−3.38 (0.00)
01/01/15	31/03/15	0.07	−0.91 (0.36)	−0.63 (0.40)	−0.20 (0.45)	−0.06 (0.48)	0.06 (0.50)
01/04/15	30/06/15	−0.61	−0.72 (0.48)	0.01 (0.63)	−0.36 (0.55)	0.86 (0.79)	−0.86 (0.46)
01/07/15	30/09/15	4.16	−0.59 (0.03)	−0.54 (0.05)	−0.87 (0.01)	−2.71 (0.00)	2.71 (0.23)
01/10/15	31/12/15	1.55	4.23 (0.87)	3.71 (0.78)	−4.54 (0.01)	1.14 (0.43)	−1.14 (0.13)
01/01/16	31/03/16	2.51	−0.17 (0.08)	0.11 (0.13)	−0.70 (0.04)	−0.41 (0.07)	0.41 (0.16)
06/05/14	24/06/14	0.16	3.89 (0.91)	4.04 (0.92)	1.70 (0.72)	−0.39 (0.39)	0.32 (0.51)

Notes: The p -values of H_0 $\$SR_{\text{sent}} SR_{\text{other}}$ are added between parentheses, SR expresses the daily Sharpe. These p -values indicate the likeliness that an alternative strategy generates higher Sharpe ratios than the sentiment strategy. For example, for the first trimester of 2016, the interday strategy using Loughran–McDonald dictionary has a Sharpe ratio of 2.51. The second highest Sharpe (0.41) for that period is generated by the up and down persistency strategy

This strategy works especially well when there is a steady decline or incline in the price. Statistical tests, similar to the ones discussed before, have been added. As one might dispute shorting on the Bitstamp exchange, I have added a no-short strategy. Not being able to short highly affects the performance of this strategy.

3.1.2 Discussion. Comparison of the annualized compounded returns and the Sharpe ratios of strategies using different dictionaries shows that there is no evidence, and that one of these strategies is a better predictor of Bitcoin price movements.

Although prediction scenarios are not as rigorous as a statistical analysis, they can help us with analyzing and better understanding the market. In this case, simulations on real historic data seem to indicate that there is a value in assessing expert media sentiment and leveraging these insights in an automated trading fashion. This would point to the Bitcoin market not being semi-strongly efficient. A factor that is often neglected, although it might explain this seeming market anomaly, is transaction costs. The trading strategy presented in [Garcia and Schweitzer's study \(2015\)](#) is shown not to be profitable with a 0.25 per cent transaction fee. [Shah and Zhang \(2014\)](#) did not mention transaction costs, though with 2,872 trades in 50 days, it is hard to believe that any profit will remain. The Bitstamp fees vary with your monthly exchange volume[18]. To both be optimistic and conservative, I assume it to fall in the tier with the lowest and the highest fee, that is 0.10 and 0.25 per cent, respectively. While the market might appear to be very inefficient at first, this seems to be largely caused by another market friction, that is transaction costs. As a result of these costs, profitability seems to be more in line with the inherent risk of Bitcoin and the complexity of this strategy ([Table IX](#)).

Start	End	Annualized compounded return			Sharpe ratio		
		Harvard	Loughran– McDonald	<i>p</i> -value	Harvard	Loughran– McDonald	<i>p</i> -value
01/01/14	31/12/15	0.45	0.82	0.6943	1.14	1.84	0.8092
01/01/14	31/12/14	0.70	1.28	0.6735	1.49	2.23	0.7378
01/01/15	31/12/15	0.10	0.32	0.6245	0.45	1.03	0.6974
01/01/14	30/06/14	0.60	0.77	0.5218	1.14	1.47	0.5838
01/07/14	31/12/14	0.99	2.06	0.7432	2.85	3.31	0.5952
01/01/15	30/06/15	−0.26	−0.15	0.5623	−0.58	−0.28	0.5729
01/07/15	31/12/15	0.87	1.10	0.5538	1.95	2.62	0.6659
01/01/14	31/03/14	−0.60	1.20	0.8452	−1.27	1.99	0.9289
01/04/14	30/06/14	3.38	1.94	0.4001	3.08	2.39	0.3843
01/07/14	30/09/14	0.51	−0.40	0.0569	1.94	−3.14	0.0105
01/10/14	31/12/14	0.88	9.45	0.9306	2.61	5.46	0.8393
01/01/15	31/03/15	−0.67	−0.07	0.7583	−2.00	0.07	0.8184
01/04/15	30/06/15	0.23	−0.17	0.3055	1.03	−0.61	0.2507
01/07/15	30/09/15	0.03	1.85	0.9067	0.22	4.16	0.9455
01/10/15	31/12/15	1.55	0.59	0.3329	2.46	1.55	0.3405
01/01/16	31/03/16	−0.36	1.71	0.8900	−1.37	2.51	0.9220
06/05/14	24/06/14	0.20	0.01	0.4343	0.66	0.16	0.4335

Notes: *P*-values are from Welch's test comparing mean return and mean Sharpe ratio of both strategies. High *p*-values suggest that we cannot conclude that strategy using Loughran–McDonald's dictionary is superior to the strategy using Harvard's dictionary and vice-versa

Table IX.
Comparison of
returns and sharpe
ratios when using
different dictionaries
for lexicon-based
sentiment analysis
and interday trading

Table X.
Returns and sharpe
ratios of trading
strategy using
Harvard's dictionary
after considering
impact of transaction
costs

Start	End	Annualized compounded return			Sharpe ratio		
		TC = 0%	TC = 0.10%	TC = 0.25%	TC = 0%	TC = 0.10%	TC = 0.25%
01/01/14	31/12/15	0.45	0.20	-0.11	1.14	0.65	-0.08
01/01/14	31/12/14	0.70	0.42	0.08	1.49	1.05	0.39
01/01/15	31/12/15	0.10	-0.10	-0.34	0.45	-0.11	-0.95
01/01/14	30/06/14	0.60	0.29	-0.07	1.14	0.74	0.14
01/07/14	31/12/14	0.99	0.71	0.35	2.85	2.25	1.32
01/01/15	30/06/15	-0.26	-0.39	-0.54	-0.58	-1.07	-1.81
01/07/15	31/12/15	0.87	0.51	0.10	1.95	1.34	0.44
01/01/14	31/03/14	-0.60	-0.68	-0.77	-1.27	-1.65	-2.22
01/04/14	30/06/14	3.38	2.51	1.52	3.08	2.66	2.03
01/07/14	30/09/14	0.51	0.29	0.02	1.94	1.24	0.19
01/10/14	31/12/14	0.88	0.61	0.27	2.61	2.00	1.06
01/01/15	31/03/15	-0.67	-0.73	-0.79	-2.00	-2.38	-2.95
01/04/15	30/06/15	0.23	0.03	-0.22	1.03	0.22	-0.97
01/07/15	30/09/15	0.03	-0.17	-0.39	0.22	-0.68	-2.01
01/10/15	31/12/15	1.55	1.05	0.47	2.46	1.93	1.12
01/01/16	31/03/16	-0.36	-0.46	-0.59	-1.37	-1.98	-2.88
06/05/14	24/06/14	0.20	-0.02	-0.28	0.66	0.13	-0.66

Table XI.
Returns and Sharpe
ratios of trading
strategy using
Loughran-
McDonald's
dictionary after
considering impact of
transaction costs

Start	End	Annualized compounded return			Sharpe ratio		
		TC = 0%	TC = 0.10%	TC = 0.25%	TC = 0%	TC = 0.10%	TC = 0.25%
01/01/14	31/12/15	0.82	0.47	0.08	1.84	1.26	0.38
01/01/14	31/12/14	1.28	0.85	0.35	2.23	1.71	0.94
01/01/15	31/12/15	0.32	0.07	-0.22	1.03	0.38	-0.60
01/01/14	30/06/14	0.77	0.42	0.02	1.47	0.99	0.27
01/07/14	31/12/14	2.06	1.51	0.87	3.31	2.76	1.93
01/01/15	30/06/15	-0.15	-0.31	-0.50	-0.28	-0.90	-1.82
01/07/15	31/12/15	1.10	0.70	0.23	2.62	1.91	0.84
01/01/14	31/03/14	1.20	0.74	0.22	1.99	1.46	0.67
01/04/14	30/06/14	1.94	1.39	0.74	2.39	1.98	1.35
01/07/14	30/09/14	-0.40	-0.48	-0.58	-3.14	-3.98	-5.19
01/10/14	31/12/14	9.45	7.15	4.61	5.46	4.90	4.06
01/01/15	31/03/15	-0.07	-0.22	-0.40	0.07	-0.34	-0.94
01/04/15	30/06/15	-0.17	-0.35	-0.56	-0.61	-1.62	-3.11
01/07/15	30/09/15	1.85	1.30	0.67	4.16	3.34	2.10
01/10/15	31/12/15	0.59	0.28	-0.09	1.55	0.89	-0.10
01/01/16	31/03/16	1.71	1.03	0.32	2.51	1.85	0.86
06/05/14	24/06/14	0.01	-0.15	-0.34	0.16	-0.53	-1.55

3.2 Limitations and improvements

The theoretical framework of this paper is solely able to capture the interaction between Bitcoin-related media content and changes in the Bitcoin price. The future research should focus on causality to identify whether media coverage has actual impact on Bitcoin price. This could be done together with media sentiment analysis for news outlets in different languages because not all Bitcoin traders are reading English media. Even in the absence of transaction costs, exploiting these inefficiencies would run into limitations preventing this

system to become a real money machine. First of all, no matter how profitable a system can be on historic data, there are no guarantees for real-time traders. Second, there is the Bitcoin liquidity that would set an upper bound on how far you could scale this strategy, before moving the market yourself, and, by doing so, eliminating the inefficiency. While computational scalability is a concern for many authors writing about automated trading, this is not the case for the algorithm proposed in this paper, which can perfectly run on a standalone PC. Also, the computation of the sentiment score is map reducible, e.g. computations per media source can be parallelized. A third limitation is the inherent risk of Bitcoin as a fairly new technology and financial asset. As a consequence, a trader might require an additional risk premium on top of the one to compensate the volatility of the profits. This is because he believes that, for example, ash crashes or other black swan events that can wipe out all profits ever made are more likely to occur for Bitcoin than for other financial assets. Something to keep in mind is that FinTech is a highly innovative space and thus over time new words will be introduced. Given that I used a large enough sample and only extracted a list of the most common words, I do not believe my results to be biased by the dictionary. Though, it is important to continuously add words to the dictionary to ensure it is not out of date.

The first improvement could be to correlate the profits with market volatility and the number of transactions and articles published and to see under what conditions our algorithm functions best. For example, a trader might be advised to not apply this strategy when volatility is getting too high to increase his/her overall profitability. Future research can continue along the same lines by investigating the predictive power and the influence of other media on Bitcoin prices, e.g. Reddit. Also, other time periods can be investigated, for example, if future maturation of the Bitcoin market would make any current inefficiency disappear. One could look for price patterns other than on the semi-short horizon. Or, one could test whether there is truly no propagation effect between the expert and the mainstream media where the former publishes the news first, causing a price movement, and the latter follows, causing a second price movement.

4. Conclusion

Our work involved measuring the influence of Bitcoin news media on investor sentiment in an attempt to better understand and predict the Bitcoin price. Data were collected for these as far back in time as possible. Natural language processing techniques were used to preprocess the data, which was then fed to our sentiment analyzer. A lexicon-based approach was taken, using a generic and finance-specific psychosocial dictionaries. Given the ability to automatically assess the polarity of an article or post, we are able to observe sentiment changes over time. By analyzing reaction patterns, which show the price movements following noteworthy news stories, I propose the interday trading strategy to maximize the expected return based on the interday price patterns. 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. Added liquidity by traders that were slow to react causes the price to drift in the sentiment direction again, until it incorporates all information. This strategy is put to the test using prediction scenarios for different time periods. Both the return and the Sharpe ratio were calculated and benchmarked against simple, random and technical strategies, market indexes and previous research. I show that for most simulated time windows, our sentiment-driven strategy would have outperformed the stock market indexes and other Bitcoin strategies. However, compared

to other research our obtained results are not the best. Though, in comparison with these systems, we trade relatively less. This is beneficial as for the transaction costs, the number of trades matters. Any exceptional return that our strategy might yield might quickly disappear after taking into account the entry-level transaction fee. However, if the trading volume is high enough, the applicable fee becomes lower. In case the lowest fee applies, results would be rather good, though not exceptional. Also remember that it is hard to prove whether there is a causal media and interday price relationship or whether it is a third factor causing both.

The main findings of this study are that with the expert media, one can predict semi-short-term Bitcoin price movements, the market initially overreacts resulting in multiple corrections and a trader who fully exploits all price movements cannot achieve abnormal returns. In doing so, he/she is hindered by the transaction costs and the elevated risk of the proposed strategy. These market frictions are the more likely explanatory factors for this seeming price inefficiency, rather than it being an anomaly by itself. The results suggest the Bitcoin price to satisfy semi-strong form of market efficiency hypothesis, which is an indication of a mature market.

Notes

1. The General Inquirer is a quantitative content analysis program that uses, among others, the Harvard psychosocial dictionary.
2. The Loughran–McDonald dictionary consists of six word lists: Fin-Pos, Fin-Neg, Fin-Unc, Fin-Lit, MW-Strong and MW-Weak. Whenever we refer to the Loughran-McDonald dictionary in this paper, we point to the Fin-Pos and the Fin-Neg word lists.
3. Available at: www.crummy.com/software/BeautifulSoup/ (accessed 6 May 2016).
4. Bitcoin(s), blockchain(s), block chain(s), cryptocurrency, cryptocurrencies, crypto currency, crypto currencies, altcoin(s), digital currency, digital currencies, virtual currency, virtual currencies.
5. Available at: <http://api.bitcoincharts.com/v1/csv/bitstampUSD.csv.gz> and www.quandl.com/data/BITSTAMP/USD-USD-BITCOIN-Exchange-Rate (accessed 6 May 2016).
6. Natural Language Toolkit (www.nltk.org [accessed 6 May 2016]).
7. A total of 176 words with modifiers: the percentage to be added to or subtracted from the term weight.
8. No, not, none, neither, never, nobody.
9. When a run does not finish before the end of the year, it is dropped.
10. Sentiment strategy from now onward will be used to indicate that the execution of this strategy involves measuring sentiment, as opposed to simple, random, technical and other strategies.
11. In practice, one would use 252, the average number of trading days on the NASDAQ and NYSE. However, Bitcoin can be traded every day of the year, so 365 is used.
12. The three-month US Treasury Bill is an often used as a proxy for the risk-free return, and this rate has been very close to zero for the past few years (www.treasury.gov/resource-center/data-chart-center/interest-rates/ [accessed 6 May 2016]).
13. Analysis: around 70 per cent of Bitcoins unspent for six months or more (www.coindesk.com/analysis-around-70-bitcoins-dormant-least-six-months/ [accessed 28 May 2017]).

14. Available at: www.investopedia.com/ask/answers/010815/what-good-sharpe-ratio.asp (accessed 6 May 2016).
15. Raw data obtained from <http://us.spindices.com> (accessed 6 May 2016).
16. Covariance is assumed to be zero, given the low fees.
17. Not independent and identically distributed.
18. Available at: www.bitstamp.net/fee_schedule/ (accessed 6 May 2016).

References

- Alstyne, M.V. (2014), "Why bitcoin has value", *Communications of the ACM*, Vol. 57 No. 5, pp. 30-32, available at: <https://cacm.acm.org/magazines/2014/5/174354-why-bitcoin-has-value/abstract>
- Brooke, J. (2009), A semantic approach to automated text sentiment analysis.
- Cambria, E., Schuller, B., Xia, Y. and Havasi, C. (2013), "New avenues in opinion mining and sentiment analysis", *IEEE Intelligent Systems*, Vol. 28 No. 2, pp. 15-21.
- Engelberg, E.J. and Parsons, A.C. (2011), "The causal impact of media in financial markets", Vol. 66 No. 1, pp. 67-97.
- Garcia, D., Juan Tessone, C., Mavrodiev, P. and Perony, N. (2014), "The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy", ArXiv e-prints, available at: <http://adsabs.harvard.edu/abs/2014arXiv1408.1494G>
- Garcia, D. and Schweitzer, F. (2015), "Social signals and algorithmic trading of bitcoin", CoRR abs/1506.01513, available at: <http://arxiv.org/abs/1506.01513>
- Grayson, M.R., Kwak, M. and Choi, A. (2014), "Using time-series and sentiment analysis to detect the determinants of bitcoin prices", *Issues in Information Systems*, Vol. 15 No. 2, pp. 350-358.
- Hayes, A. (2016), "Cryptocurrency value formation: an empirical analysis leading to a cost of production model for valuing bitcoin", *Telematics and Informatics*, available at: www.researchgate.net/publication/303094852_Cryptocurrency_Value_Formation_An_empirical_study_leading_to_a_cost_of_production_model_for_valuing_Bitcoin
- Jobson, J.D. and Korkie, B.M. (1981), "Performance hypothesis testing with the Sharpe and Treynor measures", *The Journal of Finance*, Vol. 36 No. 4, pp. 889-908, available at: www.jstor.org/stable/2327554
- Jurek, A., Mulvenna, M.D. and Bi, Y. (2015), "Improved lexicon-based sentiment analysis for social media analytics", *Security Informatics*, Vol. 4 No. 1, pp. 1-13, available at: <http://dx.doi.org/10.1186/s13388-015-0024-x>
- Kaminski, J. and Gloor, P.A. (2014), "Nowcasting the bitcoin market with twitter signals", CoRR abs/1406.7577, available at: <http://arxiv.org/abs/1406.7577>
- Kaushik, C. and Mishra, A. (2014), "A scalable, lexicon based technique for sentiment analysis", CoRR abs/1410.2265, available at: <http://arxiv.org/abs/1410.2265>
- Kristoufek, L. (2013), "Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era", *Scientific Reports* 3, 3415 EP, available at: <http://dx.doi.org/10.1038/srep03415>
- Loughran, T. and McDonald, B. (2011), "When is a liability not a liability? Textual analysis, dictionaries, and 10-ks", *The Journal of Finance*, Vol. 66 No. 1, pp. 35-65, available at: <http://dx.doi.org/10.1111/j.1540-6261.2010.01625.x>
- Medhat, W., Hassan, A. and Korashy, H. (2014), "Sentiment analysis algorithms and applications: a survey", *Ain Shams Engineering Journal*, Vol. 5 No. 4, pp. 1093-1113, available at: www.sciencedirect.com/science/article/pii/S2090447914000550

-
- Opdyke, J. (2006), "Comparing sharpe ratios: so where are the p-values?", *Journal of Asset Management*, Vol. 8 No. 5, pp. 308-336, available at: <http://ssrn.com/paper=886728>
- Pang, B., Lee, L. and Vaithyanathan, S. (2002), "Thumbs up? sentiment classification using machine learning techniques", *Proceedings of EMNLP*, pp. 79-86.
- Polasik, M., Piotrowska, A.I., Wisniewski, T.P., Kotkowski, R. and Lightfoot, G. (2015), "Price fluctuations and the use of bitcoin: an empirical inquiry", *International Journal of Electronic Commerce*, Vol. 20 No. 1, pp. 9-49.
- Serrano-Guerrero, J., Olivas, J.A., Romero, F.P. and Herrera-Viedma, E. (2015), "Sentiment analysis: a review and comparative analysis of web services", *Information Sciences*, Vol. 311, pp. 18-38. available at: www.sciencedirect.com/science/article/pii/S0020025515002054
- Shah, D. and Zhang, K. (2014), "Bayesian regression and bitcoin", CoRR abs/1410.1231, available at: <http://arxiv.org/abs/1410.1231>
- Sharpe, W.F. (1994), "The sharpe ratio", *The Journal of Portfolio Management*, Vol. 21 No. 1, p. 4958.
- Siering, M. (2012), "boom" or "ruin"-does it make a difference? using text mining and sentiment analysis to support intraday investment decisions", *Proceedings on 45th Hawaii International International Conference on Systems Science (HICSS-45 2012)*, 4-7 January, Grand Wailea, Maui, HI, IEEE Computer Society, pp. 1050-1059, available at: <http://dx.doi.org/10.1109/HICSS.2012.2>
- Soroka, S., Young, L. and Balmas, M. (2015), "Bad news or mad news? sentiment scoring of negativity, fear, and anger in news content", *The ANNALS of the American Academy of Political and Social Science*, Vol. 659 No. 1, pp. 108-121, available at: <http://ann.sagepub.com/content/659/1/108.abstract>
- Tetlock, P.C. (2007), "Giving content to investor sentiment: the role of media in the stock market", *Journal of Finance*, Vol. 62 No. 3, pp. 1139-1168, available at: <http://EconPapers.repec.org/RePEc:bla:jfinan:v:62:y:2007:i:3:p:1139-1168>
- Tetlock, P.C., Saar-tsechansky, M., Macskassy, S. (2007), "More than words: quantifying language to measure firms fundamentals".
- Yermack, D. (2013), "Is bitcoin a real currency? an economic appraisal", NBER Working Paper No. 19747, available at: www.nber.org/papers/w19747

Table AI.

Detail on the article database: number of articles or posts, number selected, total number of words and average number of words per article or post

Sources	Entries	Selected	Words	Average words
Expert news media	15,854	15,850	8,604,567	542.87
CoinDesk	5,322	5,318	3,568,592	671.04
Cointelegraph	4,601	4,601	2,995,280	651.01
NewsBTC	5,931	5,931	2,040,695	344.07

Corresponding author

Vytautas Karalevicius can be contacted at: vytautas@karalevicius.lt