How Free is the Bitcoin Market? An Empirical Analysis of Bitcoin and Media Sentiment Herding

K. C. Enevoldsen *(201607559)* & D. E. Jensen *(201609870)*

Supervisor: Riccardo Fusaroli
Social and Cultural dynamics in cognition, Cognitive Science
School of Communication and Culture, University of Aarhus
Jens Chr. Skous Vej 2, 8000 Aarhus, Denmark

Github repository: https://github.com/KennethEnevoldsen/SocialCulturalDynamics

---

Regarding Initials used in the Paper
KCE: K. C. Enevoldsen
DEJ: D. E. Jensen
We regard this paper as a joint effort, however we have tried to indicate who contributed the majority of text for each section. When initials are comma separated, the first person mentioned is the primary contributor for the following section, although the whole paper has been heavily edited by both authors. Where initials are separated by an ampersand (&), both authors contributed evenly.

Abstract (DEJ & KCE)

The cryptographic currency Bitcoin has caused major debate, some considering it the first free-market solution to online commerce (Malayeri, 2018) while others see it as a bubble fuelled by a group of young tech enthusiasts (Krugman, 2013). This study seeks to examine this discrepancy by aiming to answer whether the price of Bitcoin is influenced by herding behavior and media sentiment. This was measured using Cross-Sectional Absolute Returns and sentiment extracted from articles by CoinDesk, The New York Times, and CNBC. Applying a Bayesian analysis, we found little to no evidence of herding nor sentiment influences on the daily price of Bitcoin. Similarly, the price of Bitcoin was not found to influence the media sentiment nor our measure of herding. This may indeed indicate that the Bitcoin market operates under free-market principles. However, due to methodological limitations, one should be cautious when interpreting these results.

Keywords: *Bitcoin, Cryptocurrency, Free-market, behavioral Economics, Conformity, Herding, Group-think, Sentiment Analysis, Opinion Mining*

1.0 Introduction (KCE, DEJ)

In 2008, Satoshi Nakamoto released his infamous paper introducing Blockchain technology and Bitcoin, a digital cryptographic currency (henceforth cryptocurrency) in which he argues against the current "trust-based model" of online commerce, which uses banks as trusted mediators of trade. In January 2009, Bitcoin went online and has since seen extreme growth with 1,369% return-on-investment in 2017, followed by a vast variety of different cryptocurrencies (Kondor, Pósfai, Csabai, & Vattay, 2014). As it was originally intended, Iavorschi (2013) shows that Bitcoin is indeed in compliance with free market principles. However, free markets have been shown to succumb to group biases such as herding (Chang, Cheng, & Khorana, 2000; Chiang & Zheng, 2010). Herd behavior has been shown to decrease overall group performance (Asch, 1951; Lorenz, Rauhut, Schweitzer, & Helbing, 2011; Bang & Frith, 2017), increase market volatility (Tan, Chiang, Mason, & Nelling, 2008), and influence market returns (Ouarda, El Bouri, & Bernard, 2013). Herding tendencies, especially, have been found in emerging markets (Tan et al., 2008; Lao, & Singh, 2011), but have also been found in established markets during crisis (Chiang & Zheng, 2010).

In the current study, we seek to examine whether Bitcoin is indeed a free, well-functioning market or if it is influenced by herding the same way emerging markets have been. We wish to draw the connection that if Bitcoin indeed is heavily subjected to herding, this would undermine the working principles of free-markets on which it was intended to operate.

2.0 Literature Review

2.1 The Free Market (DEJ, KCE)

Free-market economic theories generally assume that agents interacting in an economic ecosystem are self-interested and possess a natural rationality, hereby assuming humans to be homo economicus (Berggren, 2012). Given this, markets should be self-maintaining and self-regulating. The rationale, as argued by Adam Smith (1776), is that markets have the advantage of the diversification of its members. The unrestricted access on market entrance brings different opinions towards risk, time horizons, investing styles, and access to information (Surowiecki, 2004), something minimally achieved with top-down regulation of a few members.

In 1979, Kahneman and Tversky introduced what has become known as behavioral economics, which takes into account psychological dimensions of economic decision making to reveal how humans make systematic errors in judgment. This idea of irrational behavior among investors is not new. As von Mises (1966) argues, investors are "human being[s], not an abstract notion or a mythical collective entity […] it would be nothing short of idiocy to assume that they are omniscient and infallible" (p 696). However, Surowiecki (2004) argues that even "imperfect markets populated by imperfect people could still produce near-ideal results" (p. 106).

This notion is also supported by research on group performance that indicates the aggregate group response leads to a better performance in a variety of tasks, such as estimating the weight of an ox (Galton, 1907), improving accuracy of cancer diagnosis (Kurvers et al., 2016) and other problem-solving tasks (Lorenz, Rauhut, Schweitzer, & Helbing, 2011). This phenomenon, known as 'the wisdom of the crowd' (Surowiecki, 2004), is thought to emerge due to the diversity of information in a group surpassing that of individuals and even experts, with any inherent noise canceled out by aggregation. As such the wisdom of the crowd is reliant on and improved by group diversity. Page (2014) even argues for additional benefits of group diversity such as making societies and markets more robust.

2.2 Herd Behavior (DEJ, KCE)

However, the advantages of group diversity break down when subject to biases such as herding. As described by Raafat, Chater, and Frith (2009), “[h]erding is a form of convergent social behavior that can be broadly defined as the alignment of the thoughts or behaviors of individuals in a group (herd)”.

There are a few instances in the past where herding has been known to lead to overall worse group performance such as in experiments like Milgram’s (1963) controversial electric shock experiment and Asch’s (1951) conformity experiment, which are prime examples of how group convergence can lead to unethical decision making and wrongful judgement. Moreover, research by Lorenz et al. (2011) shows that the wisdom-of-the-crowd effect can be diluted if participants are given access to others’ responses over an extended period of time. When individuals rely on the same information, the initial diversity which lead the wisdom of the crowd is removed, leading to overall worse performance (Lorenz et al., 2011). This specific kind of herd behavior, an information cascade, differs on the critical trait that individuals ignore private information when making a decision and strictly rely on shared knowledge, where herd behavior occurs when a critical mass of individuals display identical behavior, whether based on private or public information (Çelen & Kariv, 2003). This may be especially pronounced by the individual investor, who tends to be less informed than the institutional investor and therefore bases his decisions on others (Ouarda, El Bouri, & Bernard, 2013). Similarly, “when individuals experience information overload, it can impair their judgment by causing them to limit their information search and use simple heuristics instead” (Agnew, & Szykman, 2010, p. 25).

Herding phenomena within markets are realized as investors buying similarly comprised portfolios or similar investment strategies (Scharfstein & Stein, 1990; Surowiecki, 2005; Ouarda et al., 2013). For example, on October 27, 1997, the Dow Jones Industrial Average (DJIA) dropped an impressive 7.18%. According to the SEC report, the turbulence lead to the New York Stock Exchange to be shut-down for 30 minutes in a futile attempt to stop the chaos. At one point, prices were falling 0.1% a minute. This cascade caught on to other stock exchanges, including Hong Kong and London, and resulted in the financial crash of the New York Stock Exchange (Mandelbrot & Hudson, 2004). According to the practiced theories at the time, this event was not only inexplicable, but improbable beyond belief. The 1997 market crash is a prime example of the potential negative outcome of extreme herd behavior within a market.

In a similar vein, market bubbles (such as the one leading up to the 1997 crash) occur when the market continuously overestimates the value of stocks, with the price eventually correcting itself in a sudden movement (Mandelbrot & Hudson, 2004). Bubbles originate when independence, diversity and private judgment in a market dissipates, argues Surowiecki (2004). Besides economic bubbles, herding has been shown to be correlated with market volatility (Tan et al., 2008), and influencing market returns (Ouarda et al., 2013).

While herding seems irrational and generally negative, Ouarda et al. (2013) argue that rational herding can lead to improved personal performance; an investor might realize his own poor performance and shortcomings and rather follow others who possess more reliable information or competence. Herding has also been shown to lead to the spreading of new ideas and economic market innovations (Raafat et al., 2009).

To investigate herding, Chang et al. (2000) proposed dispersion, operationalized as the Cross-Sectional Absolute Deviation of returns (CSAD), to capture the clustering of returns of one asset to another or to a market. This provides a gauge of heterogeneity at the aggregate level. Chang et al. further argue that since a diverse market entails a spectrum of opinion and information, market returns not diverging substantially is an indication of market participants suppressing their own intuitions and analysis and basing their investment decisions upon a common prediction.

2.3 Sentiment (KCE, DEJ)

As a complementary as well as an explanatory measure of herding within the Bitcoin market, the influence of media sentiment was examined using sentiment analysis. An approach which computationally seeks to examine or extract sentiment from a text or corpus (Pang & Lee, 2008).

Sentiment analysis has been applied to predict investor behavior within the stock market, using sentiment derived from the Wall Street Journal's business section (Tetlock, 2007), local news (Engelberg, & Parsons, 2011) and from microblogging platforms such as StockTwits.com (Piñeiro-Chousa, Vizcaíno-González, & Erez-Pico, 2017) and Twitter (Bollen, Mao, & Zeng, 2011). The performance varies from predicting positive or negative changes in price only slightly above chance levels (Piñeiro-Chousa, et al., 2017) to predicting the directionality of the change with an accuracy of 86.7% (Bollen, et al., 2011). However, it should be mentioned that Oliveira, Cortez, and Areal (2013) found "no evidence of return predictability using sentiment indicators" when exploring the effect of sentiment derived from StockTwits.

Besides predicting behavior among stock market investors, sentiment analysis has also been proved useful in predicting how well a movie will fare in box-office sales (Asur & Huberman, 2010), state and presidential elections (Jahanbakhsh & Moon, 2014) and have, furthermore, been used as a tool for assessing political biases in news (Enevoldsen & Hansen, 2017) as well as revealing a global positivity bias (Dodds et al., 2015). Consequently, sentiment analysis is a well-tested approach utilizing the availability of written information.

In this analysis we specifically examine the three news channels; 1) The New York Times (NYT), 2) The Consumer News and Business Channel (CNBC), and 3) Coindesk (CD), which were chosen to incorporate, 1) general news media, 2) business oriented news media, and 3) cryptocurrency specific news media.

2.4 Motivation and Hypotheses (DEJ & KCE)

An underlying assumption of the orthodox economic theory is that investments are made using all available information proficiently (Scharfstein & Jeremy, 1990). Thus, an ideal world would see a market in which prices adjust efficiently to new information, an idea termed the Efficient Market Hypothesis (Blasco, Corredor, & Ferreruela, 2012). Intuition inclines us to suspect that news media regarding Bitcoin might be one of the information sources that could potentially lead to herding in the market, establishing a mechanistic link between market herding and the media.

Thus, our hypothesis H_{1a} is that we predict that the daily change of Bitcoin price can be partially determined by the sentiment of news articles referring to Bitcoin and/or cryptocurrencies. We expect this effect will be enhanced by herding behavior in the analyzed media. Furthermore, we suspect the daily movements of Bitcoin price to be partially determined by herding tendencies in the cryptocurrency market as a whole and the Bitcoin market in comparison to the general stock market (H_{2a}).

Naturally, the Bitcoin price might reciprocally influence both sentiment and the degree of herding. To encapsulate these potential circular tendencies, we hypothesize (H_{1b}) that the sentiment of these news articles is partially attributed to daily movements in Bitcoin price. Similarly, we hypothesize (H_{2b}) that the degree of news herding is influenced by the daily change in price.

3.0 Materials and Methods

3.1 Cryptocurrencies (DEJ, KCE)

Historical data of all cryptocurrencies was accessed on April 30, 2018. Data included the name, date, open, high, low, and close of each cryptocurrency dating from April 28, 2013, to January 2, 2018. The daily return of all cryptocurrencies was calculated and expressed as a percentage. The returns were then standardized.

3.2 Standard and Poor’s 500 Index (DEJ, KCE)

We chose to use Standard and Poor’s 500 (S&P 500) market index as our gauge for the general U.S. market. S&P 500 includes the 500 leading U.S. companies and captures 80% of market capitalization (S&P, 2018), a practice consistent in other studies as well (e.g. Ouarda, El Bouri, & Bernard, 2013; Engelberg & Parsons, 2011). The S&P 500 historical data was accessed through Yahoo! Finance from April 9, 2010, to April 9, 2018.

Since the cryptocurrency market does not abide by regular market hours, the missing data of the S&P 500 was calculated to match values in the cryptocurrency market by using a concave function similar to Mittal and Goel (2012) by averaging the last recorded price with the next recorded price. This generates a gradual concave transition between the missing days, which is reasonable considering that the market follows a concave trend under normal circumstances (Mittal & Goel, 2012). This was done for the open and close prices. Returns were then calculated and scaled.

3.3 Media Sentiment (KCE, DEJ)

For the sentiment analysis, a total of 10,267 articles from the period April 28, 2013, to January 2, 2018, were used from CD (n = 8,859), CNBC (n = 1,000), and NYT (n = 408) using ‘bitcoin’, ‘cryptocurrency’ and ‘cryptocurrencies’ as search terms to select articles.

To analyze the sentiment of the news, we applied a bag-of-words approach in which each word is scored based on a rated dictionary. To control for the article length, we divided the net sentiment by the total number of words, hereby deriving the article sentiment score. Prior to the rating, frequently occurring words (i.e. stop words) were removed using the tidytext package (Silge & Robinson, 2016) and the snowball stopword-list (Porter, 2001). Following the rating, the sentiment was translated into standard deviations from the mean sentiment score of the newspaper, such that positive and negative scores would become relative to each newspaper. Besides making the scores easier to interpret, scaling variables also tends to improve model fitting (McElreath, 2016). While the initial intention was to explore the different effects of news source, the degree of missing articles lead to issues with model convergence. To remedy this the mean sentiment score (MSS) for each day was calculated, hereby disregarding potential differences between news sources.

The dictionary applied was the LabMT dictionary (Dodds, Harris, Kloumann, & Danforth, 2011), which contains 10,222 rated words and wherein words are rated on a continuum of 1.3-8.5. The dictionary was chosen as it is shown to compare well against other sentiment dictionaries on a wide variety of corpora (Reagan, Tivnan, Williams, Danforth, and Dodds, 2015). Similarly, this dictionary approach was chosen due to its transparency and its ease of application and has been shown to compare well with other approaches such as machine learning (Reagan et al., 2015; Enevoldsen & Hansen, 2017).

3.4 Herding (DEJ, KCE)

Our measure of herding (CSAD) derived from Chang et al. (2000), is calculated against two markets. We first calculate the herding within the entire cryptocurrency market. Secondly, we measure herding of specifically Bitcoin to that of the general US market. We used the daily Cross-Sectional Absolute Dispersion (CSAD) as our herding measure, which was calculated using the following formula:

$$CSAD_t = \sum_{i=1}^{N} \frac{|R_{i,t} - \bar{R}_t|}{\mathrm{N} - 1}$$

To observe how each cryptocurrency herds in comparison to the rest of the cryptocurrencies, we used the above formula, where $R_{i,t}$ is the return of cryptocurrency i on day t and $\bar{R}_t$ as the averaged return of all cryptocurrencies on day t. Likewise, to quantify the herding behavior of Bitcoin in comparison to the general market, we used $R_{i,t}$ as the Bitcoin return during day t, and $\bar{R}_t$ as the S&P 500 return for day t. Finally, the CSAD was calculated to quantify the herding behavior of the media. Here, $R_{i,t}$ is the mean sentiment score of news outlet i during day t, and $\bar{R}_t$ the average sentiment for day t. To account for multiple articles per day, linear models were made with a by-news random intercept, from which the residuals were derived. This method is akin to those used by Chang et al. (2000), Verousis and Voukelatos (2018), and Ouarda, Bouri, and Bernard (2013) as a measurement of herding behavior in equity markets and exchange-traded funds.

3.5 Trend analysis (KCE, DEJ)

It has been observed that individuals have a tendency to interpret and act upon patterns, even when they are truly only random fluctuations (Kahneman, Tversky, 1972). This bias, more commonly known as the clustering illusion, was seen during the aftermath of the bombing of London (Kahneman, Tversky, 1972), and is also seen in analyzes of stock market fluctuations (Gilovich, 1991). Consequently, it seems reasonable to model this bias, which was attempted using a simple estimation of trends. We defined a trend as three or more consecutive scores in either a positive or negative direction, needing at least two values in the opposite direction to constitute an end of a trend. As one might expect longer trends to be more influential, the start of a trend is coded as one and every day the trend continues it is increased by one. Trends were calculated for the MSS and for yesterday's change in price.

4.0 Models and Analysis (DEJ & KCE)

Table 1: Model formulas for models exploring investor behavior within the Bitcoin market.

Model	Formula
M_1	daily change ~ Normal(μ,σ) $\mu \sim \alpha$
M_2	daily change ~ Normal(μ,σ) $\mu \sim \alpha + \beta_y$Yesterday's change
M_3	daily change ~ Normal(μ,σ) $\mu \sim \alpha + \beta_{MSS}MSS + \beta_s MSS_{CSAD}$ $\beta_s \sim$ Normal(0, 0.2)
M_4	daily change ~ Normal(μ,σ) $\mu \sim \alpha + \beta_b Bitcoin_{CSAD} + \beta_c Cryptocurrencies_{CSAD}$

For our analysis, a variety of Bayesian models were made using the brms package in Rstudio (RStudio Team, 2015; Bürkner, 2017) exploring herding behavior in the Bitcoin market. Models M_1-M_7 explore the influence of herding and sentiment on investor behavior, operationalized as the daily change in the price of Bitcoin. Given the nature of the analysis, the daily change in the price of Bitcoin is predicted using information from the day before. The models M_1 - M_4 and M_8-M_{10} are shown in Table 1 and 2 respectively, while models M_5-M_7 are introduced in the following section as they are variations of models M_1-M_4.

M_1 and M_2 constitute baseline models, M_1 using only an intercept as a predictor, and M_2 using yesterday's change in price. M_3 uses MSS as a predictor and M_4 models the effect of CSAD of the cryptocurrency market, the media sentiment, as well as Bitcoin in relation to the general market. M_3 and M_4 respond to hypothesis H_{1a} and H_{2a} respectively with nuances specified in following models. Models M_5-M_7 incorporates and expands upon the previous four models, where M_5 includes herding measures of M_4 and well as the MSS. M_6 is a combination of models M_3-M_5 and lastly, M_7 expands upon M_6 by including an interaction term between the MSS and its trend.

Models M_8-M_{10}, explore circular tendencies within the Bitcoin market (see Table 2). M_8 and M_9 examine whether herding is influenced by yesterday's change and the sentiment of news the previous day and responds to hypothesis H_{1b} and H_{2b}. Similarly, M_{10} examines the influence of market herding and yesterday's change on the sentiment of news. In addition to these models, baseline models using only an intercept were made.

Table 2: Model formulas for models exploring circular tendencies within bitcoin market.

Model	Formula
M_8	$Bitcoin_{CSAD} \sim Normal(\mu,\sigma)$ $\mu \sim \alpha + \beta_{MSS}MSS + \beta_y Yesterday's\ change$
M_9	$Cryptocurrencies_{CSAD} \sim Normal(\mu,\sigma)$ $\mu \sim \alpha + \beta_{MSS}MSS + \beta_y Yesterday's\ change$
M_{10}	$MSS \sim Normal(\mu,\sigma)$ $\mu \sim \alpha + \beta_b Bitcoin_{CSAD} + \beta_c Cryptocurrencies_{CSAD} + \beta_y Yesterday's\ change$

To the knowledge of the authors, studies using these methods while investigating cryptocurrencies have not previously been conducted. As such only weakly informative priors were chosen using half Student-t prior with 3 degrees of freedom, corresponding to the default priors determined by the brms package (Bürkner, 2017). This prior often leads better convergence compared to alternative weakly informative priors (Bürkner, 2017). Sentiment CSAD, as the only exception, was given a restrictive normal prior with a mean of 0 and an SD of 0.2, due to issues with convergence. Furthermore, all response distributions were assumed to follow a Gaussian distribution, as it is a reasonable assumption given its maximum entropy (McElreath, 2016, p. 289-297).

As news regarding Bitcoin is not published on a daily basis, missing values of the MSS, as well as the sentiment CSAD, were calculated by applying multiple imputations by chained equation (MICE) using the R implementation (Buuren & Groothuis-Oudshoorn, 2010). The missing values were imputed five times applying the default method of predictive mean matching using the remainder of the predictive variables. As such, each model was fit to the five datasets derived from MICE, with model estimates derived from pooling the estimates of each fit. A total of 10,000 iterations was conducted for each fit using two chains of which 4,000 iterations constituted the warm-up. Lastly, due to issues with model convergence, CSAD measures were standardized.

5.0 Results (DEJ & KCE)

Tables 3 and 5 show model estimates for models M_1-M_{10}, and performance metrics for each of the models are seen in Table 3 and 6. Figures 1-7 show the posterior predictive plot for models M_1-M_7, while figures 8-10 show the posterior predictive plot for model M_8-M_{10}.

Table 3: Displaying model estimates for model M_1-M_7. Values within parenthesis display 95% credibility interval (CI).

Coefficients	M_1	M_2	M_3	M_4	M_5	M_6	M_7
Intercept	8.53 (-0.57, 17.52)	7.44 (-1.46, 16.39)	11.29 (1.63, 20.93))	8.54 (-0.46, 17.57)	11.46 1.85, 21.11)	10.35 (0.75, 19.91)	10.28 (0.25, 20.31)
Mean Sentiment Score			-12.16 (-26.68, 2.25)		-12.79 (-27.21,1.75)	-12.43 (-26.68, 2.01)	-11.96 (-27.47, 3.67)
Bitcoin CSAD				-3.81 (-12.75,5.15)	-3.88 (-12.98,5.17)	-3.44 (-12.46, 5.54)	-3.32 (-12.35, 5.72)
Cryptocurrency CSAD				-5.34 (-14.37, 3.75)	-5.54 (-14.57, 3.53)	-5.58 (-14.57, 3.46)	-5.69 (14.65, 3.34)
Sentiment CSAD			4.23 (-3.93, 14.05)		-4.02 (-3.93,13.07)	4.47 (-3.46, 13.46)	4.32 (-3.57,13.33)
Mean Sentiment Score Trend							-0.12 (-0.55, 0.31)
Yesterday's Change		0.14 (0.09, 0.18)				0.14 (0.09, 0.18)	0.14 (0.09, 0.18)
MSS:MSSTrend							2.62 (-3.06, 8.30)
Sigma	190.78 (184.42, 197.27)	190.69 (184.47, 197.24)	190.68, (184.41,197.15)	190.77 (184.46, 197.34)	190.65 (184.36, 197.17)	189.03 (182.81, 195.51)	189.06 (182.86, 195.52)

Table 4: Displaying Performance Metrics, including Watanabe–Akaike information criterion (WAIC), in sample root mean square error (RMSE) and a Bayesian R-Squared[1]. Values within parenthesis display 95% credibility interval (CI).

1 For further information see Gelman, Goodrich, Gabry, and Ali, 2017.

Model	WAIC	SD (WAIC)	WAIC Weight	RMSE	R^2
M_1	22873.35	508.94	0.18	190.6415	1.40e-30 (4.40e-34,7.81e-30)
M_2	22895.97	486.39	0.00	188.8954	0.02 (7.70e-03, 0.03)
M_3	22871.39	507.81	0.49	190.5164	2.01e-03 (6.76e-06, 7.20e-03)
M_4	22873.15	507.73	0.20	190.4929	3.03e-03 (2.49e-04, 8.71e-03)
M_5	22874.00	508.32	0.13	190.3106	5.41e-03 (9.09e-04, 1.28e-02)
M_6	22904.04	490.41	0.00	188.5998	0.02 (0.01, 0.04)
M_7	22903.81	488.05	0.00	188.5367	0.03 (0.01, 0.04)

Table 5: Displaying model estimates for model M_8-M_{10}. Values within parenthesis display 95% credibility interval (CI).

Coefficients	M_8	M_9	M_{10}
Intercept	0.00 (-0.05, 0.05)	0.02 (-0.03, 0.07)	0.24 (0.21, 0.27)
Mean Sentiment Score	-0.00 (-0.08, 0.08)	-0.07 (-0.14, 0,00)	
Bitcoin CSAD			-0.01 (-0.04, 0.02)
Cryptocurrency CSAD			0.01 (-0.02, 0.03)
Yesterday's Change	-0.00 (-0.00, 0.00)	0.00 (-0.00, 0.00)	-0.00 (-0.00, 0.00)
Sigma	1.00 (0.97, 1.04)	1.00 (0.97, 1.03)	0.63 (0.61, 0.65)

Table 6: Displaying Performance Metrics. Values within parentheses display 95% credibility interval (CI). Notice that WAIC weight is model weight taken when compared to baseline model.

Coefficients	M_8	M_9	M_{10}
WAIC Weight compared to baseline model	0.18	1.00	0.07
RMSE	1.03	1.00	0.63
R^2	0.00 (0.00, 0.01)	0.00 (0.00, 0.01)	0.00 (0.00, 0.01)

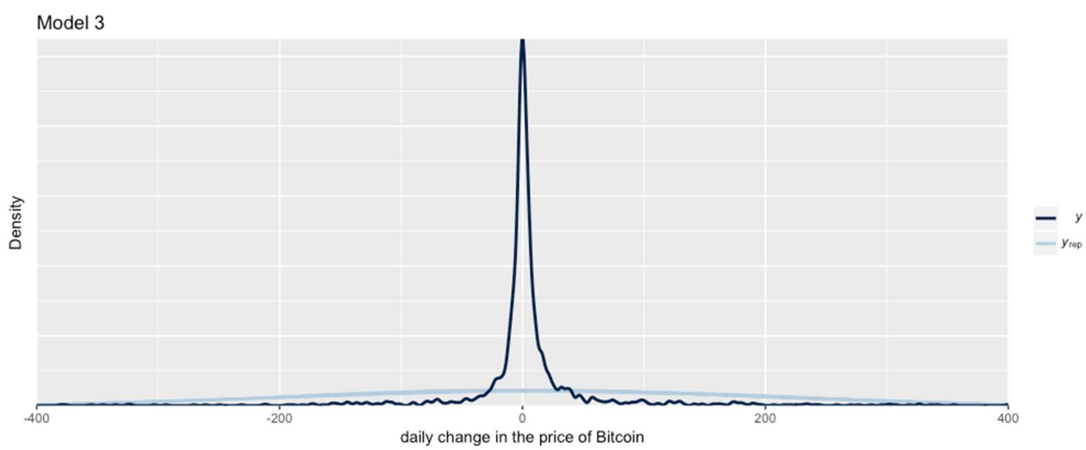

Figure 1: A posterior predictive plot derived from model, where the light blue indicates the predictive posterior using 100 samples from the model and where the dark blue indicates observed results. Notice that to improve visualization the x-axis was limited and as such exclude extreme price changes.

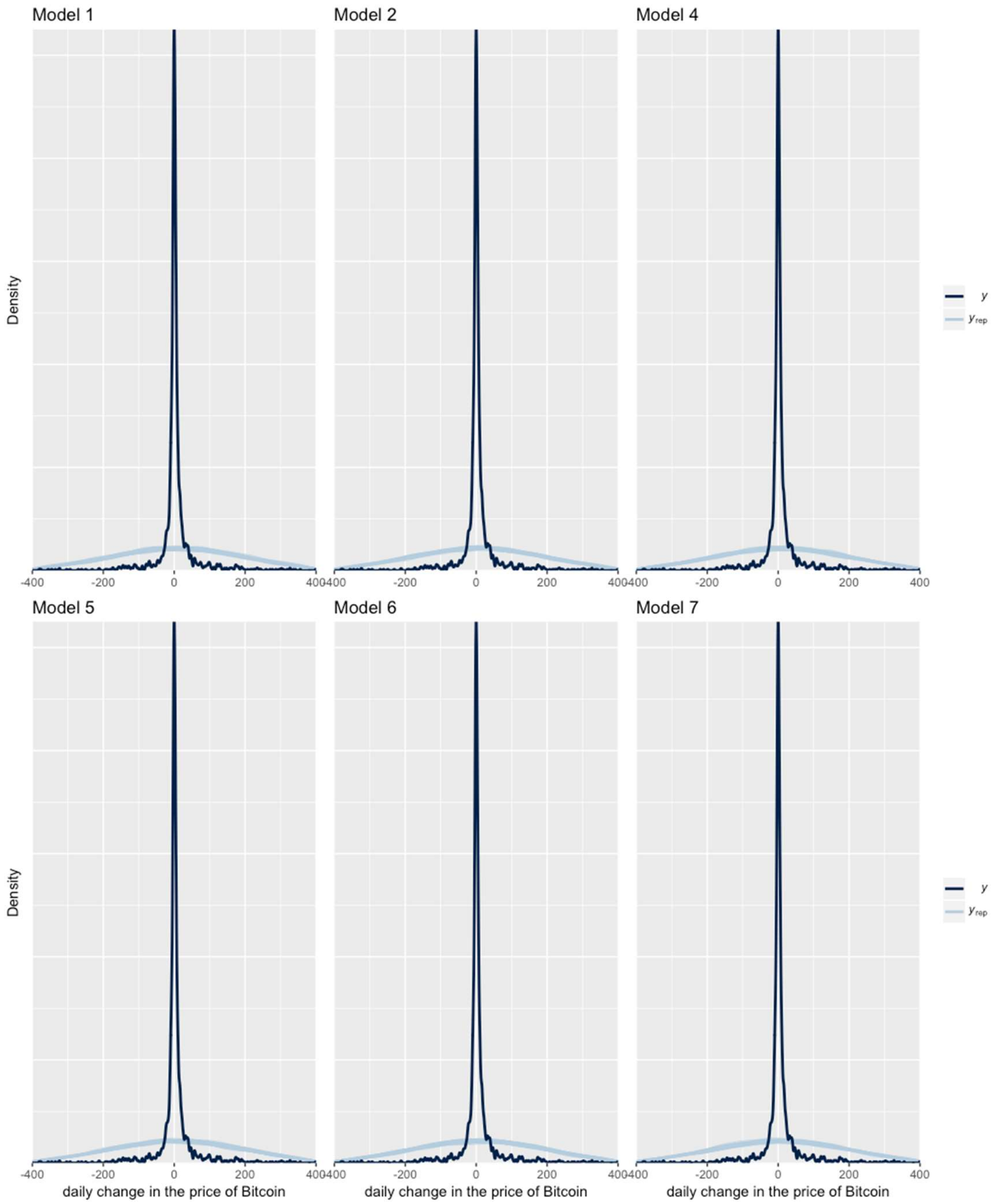

Figure 2-7: Posterior predictive plots derived from model 1, 2 and 4-7, where the light blue indicates the predictive posterior using 100 samples from the respective model and where the dark blue indicates observed results. Notice that to improve visualization the x-axis was limited and as such exclude extreme price changes.

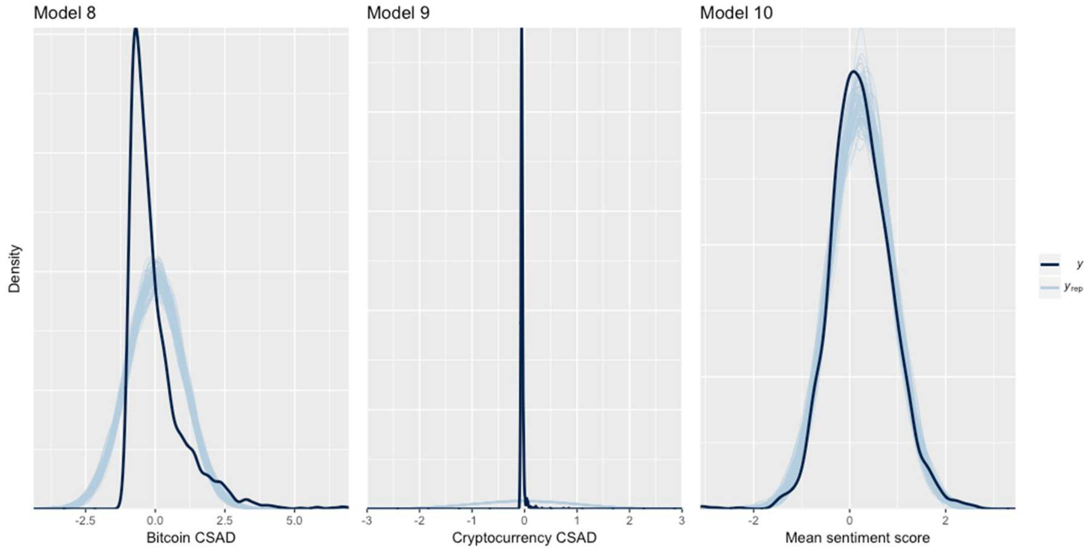

Figure 8-10: Posterior predictive plots derived from model 8-10, where the light blue indicates the predictive posterior using 100 samples from the respective model and where the dark blue indicates observed results. Notice that in figure 9 the x-axis was limited to improve visualization and as such exclude five extreme values.

6.0 Discussion

6.1 Results (DEJ & KCE)

Examining Tables 3 and 4 it can generally be seen that models M_1-M_7 performed poorly, with relatively high RMSE and very low R^2 values. Models M_3 and M_4 took most of the WAIC weights, 49% and 20% respectively. Even so, predictive posterior plots (see figures 1-7) indicate that all models fail to capture the distribution of the data. This is also inferred by the sigma estimates (see Table 3), indicating there is much uncertainty surrounding the posterior distributions, with all estimates similar to that of model M_1 with only an intercept.

A similar tendency is seen in Table 6, where it can be seen that models M_8 and M_{10} compared to a baseline model take 18% and 7% of the WAIC weight, indicating that the MSS and Bitcoin congruency towards the market is hardly predictable by the yesterday's change in price.

Model M_9 takes all of the WAIC weight when compared to a baseline model. However, examining the R^2 , RMSE (Table 6), and the sigma estimates (Table 5) we see that while the model might out-perform a baseline model, it performs poorly overall, similarly indicating that herding within the cryptocurrency market is only minutely, if at all, influenced by yesterday's change in price. This is also seen in figure 9, where the posterior predictive distribution describes the observed data poorly. While no influence of price was found on sentiment, the number of articles from NYT regarding Bitcoin was found significantly predictable from yesterday's change in price. This might indicate that the NYT might be more reactive to the Bitcoin market rather than influencing it.

Examining Table 4, it is observed that our estimated effects of models M_1-M_7 are very uncertain. Looking at the models taking most of the weights (M_3 and M_4), both sentiment and herding measures show fairly small effects with a high degree of uncertainty. Examining the RMSE and R^2 of models M_2 and M_4, we find that both have an RMSE of approximately 190 and an R^2 below 0.01, indicating that they only describe incremental variation within the data. This is supported by examining figures 1-7 in which none of the models' posterior distributions satisfactorily describes the observed data.

When examining the effect of yesterday's change, we see that there is a small but consistent effect across models $\beta = 0.14$ (0.09, 0.18), indicating our intuition that yesterday's change in price influences the following day. M_1 shows that in all ignorance, the average daily change is \$8.53 (-0.57, 17.52), whereas simply incorporating yesterday's change in model M_2 changes the intercept to \$7.44 (-1.46, 16.39). Thus, although a small effect, we see a change in estimate of over \$1 from one model to the next by including yesterday's change in price.

Besides the models mentioned above, a wide variety of variations were run to test and control for potential effects. The first one is a variation of M_2 that includes the S&P 500 to test whether it is relevant to control for a general market trend. It was found to perform significantly worse than the original model, which took almost 93.2% of the WAIC weights. Additionally, a variation of M_2 was made controlling for a trend in Bitcoin price. Similar to the MSS trend it was found to perform significantly worse compared to the original model, which took nearly 99.9% of the WAIC weights. Furthermore, models using an individual sentiment score per news source as well as a random effect per day were also tested, however, due to a high amount of missing values these models had trouble converging even when missing values were estimated using MICE. It is unknown whether including more articles might improve the estimation of missing articles sentiment and lead to an overall better fit, something which future studies should examine.

Finally, a variation of M_7 was made testing whether an interaction between the sentiment CSAD and MSS, as one might expect the news to be more influential if the media agrees with each other. The variant was found to perform significantly worse with the original model taking almost 99.7% of the WAIC weights. Similarly, a model was made testing for a potential effect of a three-way interaction between MSS, its trend, and sentiment herding, which found null results.

Contrary to intuition, these results provide little to no evidence to support the assumption that sentiment of the media has a significant influence on investors' behavior. Our models show that all things considered, predictors of daily change are small and highly uncertain, with the exception of yesterday's change in price. As our data only includes three news sources, it is hard to be certain whether an effect is present. Future research should seek to include a more representative sample of news articles. Likewise, there is little to no evidence supporting that herding influences Bitcoin's price nor that the Bitcoin price influences the sentiment of news regarding cryptocurrencies and herding within the market.

6.2 Methodological Limitations (DEJ & KCE)

Our analysis does not explore the possibility of the efficiency of the market reacting to news. Media online can be accessed any time after publication, thus, even if sentiment had a latent influence on price, we are unable to capture it with our current methods. Likewise, we remain uncertain as to the degree of efficiency the market adjusts to news. It is possible that a) the sentiment is already realized in the price, b) that the market has a slow turnover to news, or c) possibly even agnostic to news.

One might notice that previous studies examining stock market behavior have simplified the prediction into a binomial classification predicting positive or negative changes (Bollen et al., 2011; Madan, Saluja & Zhao, 2015; Piñeiro-Chousa et al., 2017). However, applying the same approach found no models outperforming the baseline model, which obtained an accuracy of 54.36% by assuming that the change in Bitcoin price would always be positive. As such, this approach seems to lead to similar conclusions as the one applied.

As already mentioned, our measure of trends was simple and might be too ill-defined to capture the nuances of trends and their long-term effects. A possible solution might be to include indicators used in technical analysis as a guidance of defining trends. Additionally, the classification of trends in the present study may lean more in the direction of short-term momentum than long-term movements. Two different investor strategies include considering short-term or long-term trends (Filbeck et al., 2017) and therefore a more robust analysis could include two thresholds that capture long-term and short-term trends.

As previously mentioned (see *4.0 Models and Analysis*), the use of MICE has obtained wide acceptance. However, Buuren and Groothuis-Oudshoorn (2010) mention that “we need a better understanding of the dangers and limitations of the technique” and, therefore results should be interpreted with caution. However, repeating the analysis with the missing values omitted leads to overall higher RMSE and lower R^2 scores, implying the approach may be a useful tool for dealing with missing values. Again, with the degree of uncertainty obtained, no definite conclusions should be derived.

While we have assumed that daily changes follow a normal distribution as traditional economics assumes, Mandelbrot and Hudson (2004) note that there are too many extreme values too often for price changes to be categorized as a normal distribution. Similarly, examining figures 8 and 9, one might argue that a Gaussian assumption is flawed, however, exploring multiple response distributions in models M_8 and M_9 did not lead to a change in inference.

6.2.1. Herding (DEJ, KCE)

The measure of herding applied in this study was adapted from previous studies (Chang et al., 2000; Ouarda et al., 2013; Verousis & Voukelatos, 2018) where it was used to capture herding in markets such as exchange-traded funds and financial markets. However, the previously mentioned markets are not based on decentralized regulation, which may make the Bitcoin market incomparable to the markets in which the measure of herding has been applied. As Bitcoin is a digital currency, it might have been more appropriate to compare it to the foreign exchange market, with the extra advantage being that the foreign exchange market reflects the Bitcoin market more in terms of regulation and trading hours than the S&P 500.

Our analysis of herding includes an increasing number of cryptocurrencies until 2014 where the number of cryptocurrencies reaches an equilibrium of births and deaths, corresponding to one new currency a week for every one abandoned (ElBahrawy, Alessandretti, Kandler, Pastor-Satorras, & Baronchelli, 2017). Whether this has any influence on the herding measure or not is uncertain. Similarly, we have assumed equal influence of the different cryptocurrencies, where it might be advantageous to weigh cryptocurrencies according to their trade volume or market share.

Moreover, our calculation of CSAD is one of many methods of measuring herding. Under the Capital Asset Pricing Model (CAPM), a widely used paradigm of trading, it is predicted that the absolute dispersion of returns is a linear function (Ouarda et al., 2013). However, as observed by Chang et al. (2000), during times of crisis and high volatility this measure becomes asymmetric. To overcome these issues, Ouarda et al. 2013 suggest the following formula, which might prove better at dealing with the high volatilities in the price of Bitcoin:

$$CSAD_t = \gamma_0 + \gamma_1 R_{m,t} + \gamma_2 \left| R_{m,t} \right| + \gamma_3 R_{m,t}^2 + \varepsilon$$

Other approaches also theorized to reflect herding include realized, historical, or implied volatility (Blasco, Corredor, & Ferreruela, 2012) and volume of transactions (Ouarda, El Bouri, & Bernard, 2013). Our current measure of herding is in relation to the rest of the cryptocurrency market and Bitcoin's congruency toward the S&P 500, when it would be more accurate if we explored herding tendencies within the Bitcoin market itself. This can be achieved using the trades conducted within the Bitcoin community or analyzing the trade volumes. All of these approaches would have to overcome the obstacle of automated trading which might dilute a potential effect of human psychological biases.

Another approach to measuring herding is on the individual level using the Lakonishok, Shleifer, and Vishny (LSV) measure (Sharma, & Bikhchandani, 2000). This is based on trades conducted on a subset of market participants over a period of time. This more direct approach might prove more efficient as the trades conducted are publicly available (Nakamoto, 2008), which makes the possibility for further examining the herding within subgroups.

6.2.2 The Sentiment Approach (KCE, DEJ)

The present study analyzes articles from three different sources: CoinDesk, The New York Times, and CNBC. CNBC specifically is known for being a primary outlet of economic and financial news, The New York Times caters to a more general audience while CoinDesk is primarily a cryptocurrency news outlet. These were selected to incorporate a wide variety of sources and to explore a potential difference in effect. However, due to issues with convergence, sentiment scores were aggregated into one daily mean sentiment, an approach which is questionable as CD included multiple articles for almost every day, while NYT and CNBC had periods with no articles regarding Bitcoin and NYT rarely had multiple articles per day. As a result, the MSS is predominantly influenced by news from CD. Furthermore, although these three sources have a certain variety, three sources may not be enough to capture the bulk of the sentiment in the news. For a more rigorous analysis, one could analyze news from a wider range of sources.

A different approach could also be to analyze commercial news sentiment versus public sentiment, derived from tweets, Facebook posts, or other forms of online data. This would be theoretically interesting to see how the public opinion diverges from that of the corporate, and possibly look into which one is more predictive of Bitcoin price changes. While our approach represents a variety of news which could influence investors, it can be argued that a sentiment analysis is better applied to microblogs with more explicit valance, similar to approaches by Mittal and Goel (2012) and Piñeiro-Chousa et al. (2017). This is especially relevant considering that news are "expected to live up to journalistic ideals of objectivity" (Enevoldsen & Hansen, 2017) and consequently may prove more effective to use sentiment within more emotionally laden media, such as Reddit, Twitter, or comment sections on popular sites.

The sentiment analysis applied in this paper makes the disputable bag-of-words assumption, which assumes the text to be a "bag of words", hereby disregarding syntax, cotext, and context. This is problematic as the context can significantly change the sentiment of words, for instance, the most positively rated word, *'laughter',* might be spiteful in certain cases. The use of simple n-grams alleviates issues such as rating '*not happy*', which would normally be rated as positive, as negative. However, to capture more complex sentiment it might be ideal to use alternate methods such as the deep learning approach applied by Glorot, Bordes, and Bengio (2011). Deep learning or similar machine learning approaches would require a rated dataset of news to be trained upon before it can be utilized for the remainder of the dataset. Similarly, machine learning methods have less transparency and it can be hard to discern the influence of certain words or phrases (Reagan et al., 2015).

6.3 Implications (DEJ & KCE)

Though our approach has methodological limitations, it incorporates sources from three different classes of news with almost all articles available since Bitcoin entered public attention. Similarly, our measure of herding incorporates all data available since this period and is an accredited method that has proven descriptive in similar cases (Chang et al., 2000; Ouarda et al., 2013; Verousis & Voukelatos, 2018). As such, if we assume that our approach will generalize despite its limitations, our analysis suggests that the Bitcoin price is hardly influenced by sentiment nor herding within the market. Furthermore, it also suggests news sentiment is minimally affected by herding in the market and the Bitcoin price, and vice versa for herding. This seems to indicate that the Bitcoin market has not fallen prey to the consequences of herding and sentiment as other markets. Therefore, the Bitcoin market may indeed functionally utilize the free-market principle of the wisdom of the crowd, being self-maintaining, and self-regulating.

Conclusion (DEJ & KCE)

As has been stated in the past, the market is nearly impossible to predict with consistent accuracy (Mandelbrot & Hudson, 2004). Our results seem to support this notion, as they provide little to no evidence to support the assertion that sentiment of the media or market herding has a significant influence on the price of Bitcoin. This indicates that the Bitcoin market operates under the free-market principle of the wisdom of the crowd allowing it to be self-maintaining and self-regulating. However, due to methodological limitations, one should be cautious when interpreting these results. Noteworthy limitations include that our herding measure might not properly measure herding in the Bitcoin community and that sentiment disregards word cotext and context. Future studies might seek to directly measure herding within Bitcoin using information available in the ledger and include cotext and context into the sentiment analysis using machine learning approaches. However, whether this will change inference remains unknown.

7.1 References

Action, H. (1966). A Treatise on Economics. *Scholar's Edition published by Ludwig von Mises*.

Agnew, J. R., & Szykman, L. R. (2005). Asset allocation and information overload: The influence of information display, asset choice, and investor experience. In Bruce, Brian. (2010). *Handbook of Behavioral Finance.* Edward Elgar Publishing Ltd. Cheltenham, United Kingdom.

Asch, S. E., & Guetzkow, H. (1951). Effects of group pressure upon the modification and distortion of judgments. *Groups, Leadership, and Men*, 222–236.

Asur, S., & Huberman, B. A. (2010). Predicting the Future With Social Media. *IEEE Computer Society.* Retrieved from http://www.hpl.hp.com/research/scl/papers/socialmedia/socialmedia.pdf?utm_source=twitterfeed&utm_medium=twitter

Bang, D., & Frith, C. D. (2017). Making better decisions in groups. *Royal Society Open Science*, *4*(8), 170193. https://doi.org/10.1098/rsos.170193

Berggren, N. (2012). Time for behavioral political economy? An analysis of articles in behavioral economics. *The Review of Austrian Economics*, *25*(3), 199-221.

Blasco, N., Corredor, P., & Ferreruela, S. (2012). Does herding affect volatility? Implications for the Spanish stock market. *Quantitative Finance*, *12*(2), 311-327.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/J.JOCS.2010.12.007

Buuren, S. van, & Groothuis-Oudshoorn, K. (2010). MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 1–68. Retrieved from https://dspace.library.uu.nl/handle/1874/44635

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Chang, E. C., Cheng, J. W., & Khorana, A. (2000). An examination of herd behavior in equity markets: An international perspective. *Journal of Banking & Finance*, *24*(10), 1651-1679.

Chiang, T. C., & Zheng, D. (2010). An empirical analysis of herd behavior in global stock markets. *Journal of Banking & Finance*, *34*(8), 1911–1921.

Çelen, B., & Kariv, S. (2004). Distinguishing informational cascades from herd behavior in the laboratory. *American Economic Review*, *94*(3), 484-498.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PloS One*, *6*(12), e26752.

Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., … Danforth, C. M. (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(8), 2389–94. https://doi.org/10.1073/pnas.1411678112

ElBahrawy, A., Alessandretti, L., Kandler, A., Pastor-Satorras, R., & Baronchelli, A. (2017). Evolutionary dynamics of the cryptocurrency market. *Royal Society open science*, *4*(11), 170623.

Enevoldsen, K. C., & Hansen, L. (2017). Analysing Political Biases in Danish Newspapers Using Sentiment Analysis. *Journal of Language Works-Sprogvidenskabeligt Studentertidsskrift*, *2*(2), 87–98.

Engelberg, J. E., & Parsons, C. A. (2011). The Causal Impact of Media in Financial Markets. *The Journal of Finance*. WileyAmerican Finance Association. https://doi.org/10.2307/29789772

Filbeck, G., Ricciardi, V., Evensky, H. R., Fan, S. Z., Holzhauer, H. M., & Spieler, A. (2017). Behavioral finance: A panel discussion. *Journal of Behavioral and Experimental Finance*, *15*, 52-58.

Galton, F. (1907). Vox populi (The wisdom of crowds). *Nature*, *75*(7), 450–451.

Gelman, A., Goodrich, B., Gabry, J., & Ali, I. (2017). R-squared for Bayesian regression models.

Glorot, X., Bordes, A., & Bengio, Y. (2011). Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th international conference on machine learning (ICML-11)* (pp. 513–520).

Grolemund, G., Wickham, H. (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL http://www.jstatsoft.org/v40/i03/.

Iavorschi, M. (2013). The bitcoin project and the free market. *CES Working Papers*, *5*(4), 529–534.

Jahanbakhsh, K., & Moon, Y. (2014). The Predictive Power of Social Media: On the Predictability of U.S. Presidential Elections using Twitter. Retrieved from http://arxiv.org/abs/1407.0622

Bitcoin Is Evil. (2013, December 28). Retrieved June 2, 2018, from https://krugman.blogs.nytimes.com/2013/12/28/bitcoin-is-evil/

Kondor, D., Pósfai, M., Csabai, I., & Vattay, G. (2014). Do the Rich Get Richer? An Empirical Analysis of the Bitcoin Transaction Network. *PLoS ONE*, *9*(2), e86197. https://doi.org/10.1371/journal.pone.0086197

Kurvers, R. H. J. M., Herzog, S. M., Hertwig, R., Krause, J., Carney, P. A., Bogart, A., … Wolf, M. (2016). Boosting medical diagnostics by pooling independent judgments. *Proceedings of the National Academy of Sciences*, *113*(31), 8777–8782.

Lao, P., & Singh, H. (2011). Herding behavior in the Chinese and Indian stock markets. *Journal of Asian Economics*, *22*(6), 495–506.

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, *108*(22), 9020–9025.

Madan, I., Saluja, S., & Zhao, A. (2015). Automated Bitcoin Trading via Machine Learning Algorithms. *URL: Http://Cs229. Stanford. Edu/Proj2014/Isaac\% 20Madan*, *20*, 1–5. Retrieved from https://pdfs.semanticscholar.org/e065/3631b4a476abf5276a264f6bbff40b132061.pdf

Malayeri, M. J. (2018). *How Hayek's Vision of Free Market Money Has Become Reality*.

https://www.austriancenter.com/hayek-vision-free-market-money-reality/

Mandelbrot, B., & Hudson, R. L. (2004). The misbehavior of markets. New York, NY: *Basic Books.*

McElreath, R. (2016). *A Review of Statistical Rethinking: A Bayesian Course With Examples in R and Stan*. *Apple Academic Press Inc.* https://doi.org/10.3102/1076998616659752

Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, *67*(4), 371.

Mittal, A., & Goel, A. (2012). Stock prediction using twitter sentiment analysis. *Standford University,* from *http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)*, *15*.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Bitcoin.Org*.

Oliveira, N., Cortez, P., & Areal, N. (2013). On the Predictability of Stock Market Behavior Using StockTwits Sentiment and Posting Volume (pp. 355–365). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40669-0_31

Ouarda, M., El Bouri, A., & Bernard, O. (2013). Herding behavior under markets condition: Empirical evidence on the European financial markets. *International Journal of Economics and Financial Issues*, *3*(1), 214.

Page, S. E. (2014). Where diversity comes from and why it matters? *European Journal of Social Psychology*, *44*(4), 267–279. https://doi.org/10.1002/ejsp.2016

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, *2*(1–2), 1–135. https://doi.org/10.1561/1500000011

Pedersen, T. L. (2017). patchwork: The Composer of ggplots. R package version 0.0.1.

https://github.com/thomasp85/patchwork

Piñeiro-Chousa, J., Vizcaíno-González, M., & Erez-Pico, A. M. (2017). Influence of Social Media over the Stock Market. *Psychology & Marketing*, *43*. https://doi.org/10.1002/mar.20976

Porter, M. F. (2001). Snowball: A language for stemming algorithms.

Raafat, R. M., Chater, N., & Frith, C. (2009). Herding in humans. *Trends in Cognitive Sciences*, *13*(10), 420–428.

Reagan, A. J., Danforth, C. M., Tivnan, B., Williams, J. R., & Dodds, P. S. (2016). Benchmarking sentiment analysis methods for large-scale texts: A case for using continuum-scored words and word shift graphs. Retrieved from https://arxiv.org/pdf/1512.00531.pdf

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Scharfstein, D. S. & Jeremy, C. S. (1990). Herd behavior and investment. *The American Economic Review*, 80(3), 465-479.

Sharma, M. S., & Bikhchandani, S. (2000). Herd Behavior in Financial Markets: A Review (No. 0-48). *International Monetary Fund.*

Silge, J., & Robinson, D. (2016). "tidytext: Text Mining and Analysis Using Tidy Data Principles in R." _JOSS_,*1*(3). doi: 10.21105/joss.00037 (URL: http://doi.org/10.21105/joss.00037), URL:http://dx.doi.org/10.21105/joss.00037>.

Smith, A. (1950). *An Inquiry into the Nature and Causes of the Wealth of Nations,(1776)*. Methuen.

Surowiecki, J. (2004). *The wisdom of crowds:* Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. *Doubleday*.

Tan, L., Chiang, T. C., Mason, J. R., & Nelling, E. (2008). Herding behavior in Chinese stock markets: An examination of A and B shares. *Pacific-Basin Finance Journal*, *16*(1–2), 61–77.

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Verousis, T., & Voukelatos, N. (2018). Cross-sectional dispersion and expected returns. *Quantitative Finance*, *18*(5), 813-826.

Wickham, H. (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL http://www.jstatsoft.org/v21/i12/.

Wickham, H., Francois, R., Henry, L., & Müller, K. (2017). dplyr: A Grammar of Data Manipulation. R package version 0.7.4. https://CRAN.R-project.org/package=dplyr

Wickham, H. (2018). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.3.0. https://CRAN.R-project.org/package=stringr

S&P Dow Jones Indices. (n.d.). Retrieved May 29, 2018, from https://ca.spindices.com/indices/equity/sp-500