On the Dynamics of Bitcoin in Relation to Social and Economic Indicators

Jack D. Hughes, [1] Alexander S. Leonardi, [2] William C. Leung, [3] Justin R. Samovar [4]

[1]San Marcos Senior High School, jdhughes@me.com, [2]Hunter College High School, alexanderleonardi@hunterschools.org [3]Saratoga High School, leuw3710@lgsstudent.org, [4]Viewpoint School, j.samovar20@viewpoint.org

Abstract

Over the past few years, cryptocurrency has gone from an esoteric concept understood only by a select few professionals to a ubiquitous, booming source of income for amateur and expert investors alike. The rise of this formerly obscure commodity can be largely attributed to the successes of Bitcoin and Ethereum, both of which have garnered enormous attention on social media and news platforms. Due to their lack of direct ties to other systems of currency (unlike many national currencies, which are backed by others), the value of Bitcoin and Etherium are notoriously difficult to predict. Research has previously indicated that cryptocurrencies' true values lie not in outside economic factors, but in their social perception. In other words, cryptocurrency value can be most heavily tied to its popularity and ubiquity in social networks. This paper seeks to investigate the extent to which this principle holds true, by analyzing the correlation between Bitcoin prices and various social and economic circumstances. In doing so, we hope to reach a comprehensive conclusion about the nature of cryptocurrency volatility.

1 Introduction

On January 1, 2017, one bitcoin was worth $1,003. By December 17 of the same year, Bitcoin had reached an all-time high of $20,089 per unit. Within three months, its value had plummeted below $8,000 per unit. This fluctuation in price exemplifies Bitcoin's volatile nature. The extent of Bitcoin's volatility is unprecedented–changing drastically and seemingly unpredictably on a regular basis. Unlike other commodities, such as gold or government-issued fiat currencies, Bitcoin's value is not determined by trust in a government entity nor is it backed by a financial reserve. Instead, a bitcoin's worth is dictated entirely by social perception. Resultantly, it seems as though the more people talk about Bitcoin, the more its value seems to rise.

This fundamental principle of cryptocurrencies is convenient for sellers hoping to profit off of the surging popularity of Bitcoin; a multitude of companies, including Microsoft, AT&T, and Expedia, accept Bitcoin as payment, with support for the platform pointing towards an upward trend. Bitcoin's unique method of operation also results in other implications that set it apart from more traditional currencies. Since transactions occur over a peer-peer network and are handled by a blockchain, the need for a third party is eliminated, allowing trading to remain anonymous. Because of this, Bitcoin has remained notorious for prevalent use on The Silk Road, a dark web marketplace infamous for selling illegal substances. Due to the dichotomy between adoption by large corporations and online use for illegal purchases, Bitcoin possesses a controversial reputation in terms of both stability and anonymity. Its mixed reputation combined with the relationship between its value and perception by the public means extreme volatility as public opinion continues to change. This paper hopes to find the answer to the following research question:

To what extent can social and economic factors demonstrate correlation with the price of Bitcoin through the implementation of a machine learning algorithm?

To answer this question, we will conduct an analysis of the correlation between a variety of features on the volatility of Bitcoin. These features will be categorized into two datasets: one consisting of social features, and the other based on economic indicators. First, we construct a multiple linear regression model, training it three times: once on the social dataset, once on the economic dataset, and a third time on both datasets combined, producing three separate price predictions. The error of each prediction with respect to the actual bitcoin price data will be evaluated, providing insight about which features are most reflected in Bitcoin's value.

2 Background and Literature Review

As the first major decentralized cryptocurrency, Bitcoin sought to change the ways in which payment processing functions in the world of e-commerce. To accomplish this, Nakamoto (2008) included key attributes in the exchange system that differentiated it from currencies of the past. First, Bitcoin transactions are made difficult to trace by their peer-to-peer nature, redefining the role of a third party in payment processing systems. Second, and more importantly, as the first cryptocurrency of its kind, economists have found it difficult to find an accurate method to predict its volatility. Theoretically, Bitcoin's value is entirely untethered to outside economic events, prompting experts to debate whether Bitcoin should be considered a currency in the traditional sense.

Yermack (2013) concluded that Bitcoin is not a currency because it faces challenges that prevent it from satisfying the criteria of conventional government-issued money. These criteria are acting as a 'medium of exchange', 'unit of account', and 'store of value.' This means that, respectively, Bitcoin fails to be widely as a means of purchasing goods and services, to conform to conventional usage of only two decimal places, and to possess a relatively stable value that allows meaningful comparison to other economic indicators and commodities. Through a detailed analysis of the circulation of bitcoins, Yermack decided Bitcoin is too volatile to be considered a currency and is, therefore, more of a speculative investment.

Huberman et al. (2018) provide further comparisons between Bitcoin and conventional currencies. After modeling the platform, their paper deduces that peer-to-peer transactions eliminate inefficiencies caused by a traditional market, at the cost of reduced stability. This paper suggests that the inherent differences between Bitcoin and other currencies ultimately reflect in the ways by which the currency functions–specifically in the context of its volatility. This volatility is speculated by conventional research to be a factor of social confirmation rather than outside economic indicators.

However, this does not completely eliminate the possibility of a link between Bitcoin and the stock market–both of which involve a degree of seemingly random volatility. Kjærland et al. (2018) conclude that Bitcoin's price is affected by both SP 500 returns and Google searches, by implementing a generalized autoregressive conditional heteroskedasticity (GARCH) model. Whereas this method fundamentally differs from the one used in this paper, the work provides confirmation that a correlation between Bitcoin and a variety of social and economic factors exists.

Mai et al. (2015) discuss how social media platforms influence Bitcoin prices, in both short and long-term contexts. Ultimately, their research indicates that any discussion of Bitcoin on social media tends to correlate with a change in value.

Additionally, Aggarwal et al. (2019) discussed the importance of social factors in the cryptocurrency market by analyzing various attributes, including news sources, government regulations, and opinion dynamics, in relation to price "peaks and drops". A lack of comprehensive data limited the paper's scope, which the authors acknowledged in their conclusion, though they suggested that the methods would have produced more meaningful results if cryptocurrencies behaved more similarly to the stock market. This demonstrates that cryptocurrencies' fundamental differences from the stock market warrant the use of a different approach towards the construction of a predictive model. In most contexts, regression-based algorithms are widely accepted to be the most accurate in predicting currency volatility, meaning this model would be preferable in the application of Bitcoin.

Chang et al. (2000) employ a multiple regression model to evaluate the effects of various economic indicators on the stock market. Their model proves to fit their data exceptionally well, with an R^2 value of 0.987. From this data, their results conclude that regression is a viable algorithm for computing correlation between economic features, confirming the validity of the objective of this paper.

3 Methods

3.1 Multiple Linear Regression

In our approach towards solving this problem, we decided to use a supervised learning model. In contrast with other machine learning methodologies such as unsupervised or reinforcement learning, this model employs the use of a large set of labeled data points to produce an output. Specifically, our approach focuses on linear regression: a subset of supervised learning that attempts to find a linear correlation between data points, generating a trendline to generalize a function that reflects the dataset.

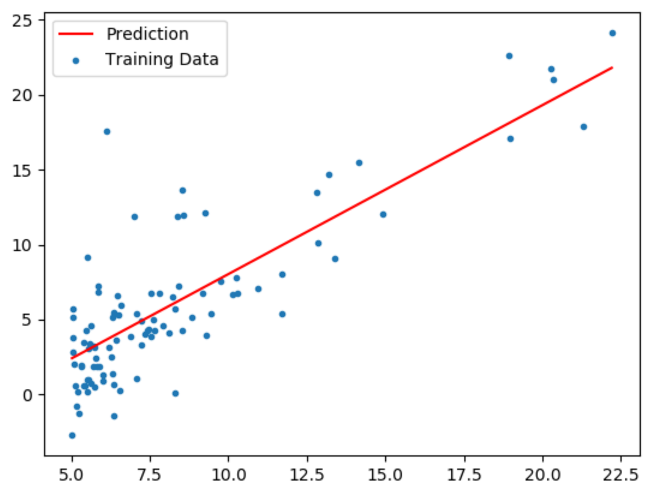

Figure 1: An example of linear regression being used to identify a linear trend in a set of points

The method portrayed in Figure 1 functions by generating a series of coefficients in the equation for a trendline $h_\theta(x)$. For the dataset depicted in the scatter plot, which contains one set of input variables x, and one set of output variables y, the regression algorithm creates two coefficients θ, and the trendline is produced using the following equation:

$$h_\theta(x) = \theta_0 + \theta_1 x \tag{1}$$

When using a one-dimensional vector of input variables, θ_0 represents its y-intercept, while θ_1 represents the trendline's slope.

For a problem as complex as currency prediction however, one set of input variables is unlikely to produce an accurate response. Therefore, we employ multiple linear regression–an algorithm that applies the concepts of linear regression to a multi-featured dataset. This model is far more capable than one-dimensional linear regression, as it can be used to predict trends that are more complex than a simple linear relation. For this method, the number of theta values must be altered to correspond with the number of features used in computation. The number of theta values t can be expressed by the following equation, where n represents the number of features:

$$t = n + 1 \tag{2}$$

Now, linear regression scales nicely for a multi-featured set of input variables, taking the form of the following equation:

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + ... + \theta_m x_m \tag{3}$$

When input variables x are stored in an $m \times n$ matrix (where m represents the number of samples), and θ is expressed as a matrix of shape $1 \times n$, the above equation can be abstracted as follows:

$$h_\theta(x) = \theta x^T \tag{4}$$

Once this computation has been performed, $h_\theta(x)$ can be evaluated and compared to a vector of given y values, or bitcoin prices. Using an optimization algorithm, θ can be adjusted to better fit the data.

3.2 Optimization

In order to ensure that the model converges on an appropriate trend, an optimization algorithm must be used. We utilized gradient descent– a method whose goal is to minimize a cost function, $J(\theta)$. In our model, the employed cost function computes mean squared error:

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{5}$$

To to minimize $J(\theta)$, gradient descent alters θ over a number of iterations by subtracting the function's partial derivative. In doing so, θ is changed by smaller increments on each iteration to prevent skipping over a global minimum. The function's learning coefficient is represented by α, which can be adjusted to change the rate at which θ is altered. If α is too high, gradient descent occurs quickly but may struggle to converge. If it is too low, the algorithm runs slower and presents a risk of over-fitting the training data.

$$\theta := \theta - \frac{\alpha}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x \tag{6}$$

This procedure is repeated until the difference between $J(\theta)$ and $J(\theta - \alpha\frac{\partial}{\partial\theta}J(\theta))$ becomes negligible. The algorithm is depicted below in pseudocode:

while $J(\theta_j) - J(\theta - \alpha\frac{\partial}{\partial\theta}J(\theta)) >$ *some small number* **do**
  $\theta_i = \theta$
  $\theta = \theta - \alpha\frac{\partial}{\partial\theta}J(\theta)$
end

4 Data

In order to accurately predict the value of a currency as volatile as Bitcoin, a dataset consisting of multiple features must be compiled. Since this study tests the impact of economic compared to social influences on the price of Bitcoin, we selected five features: daily number of tweets mentioning "Bitcoin", Google Trends data reflecting the daily number of Google searches for "Bitcoin", the S&P 500 Index, and the exchange rates between US Dollars (USD) and both Euros and Japanese Yen. A dataset consisting of daily Bitcoin prices was collected from coinmarketcap.com.

4.1 Social Data

Since 2010, Twitter has grown to over 325 million daily active users, enabling it to wield considerable influence in major events, both local and global. Consequently, Twitter is often portrayed as a reliable indicator of social trends. Due to this assumption, we chose to a dataset from bitinfocharts.com, which provides the daily number of tweets mentioning "Bitcoin".

Similarly, Google has also become ubiquitous, making it synonymous with the internet itself. Furthermore, Google search histories demonstrate useful global relationships. To collect this data, Google developed Google Trends, an API which allows users to view search data related to a specific query. We used Google Trends to search for the keyword "Bitcoin" in all categories. The values returned are integers from zero to one hundred, which compare weekly popularity of the keyword to total volume as a percent of the all-time searches. This creates some technical issues concerning the precision of our model due to weekly rather than daily values. This will be addressed later when we describe how the data was preprocessed.

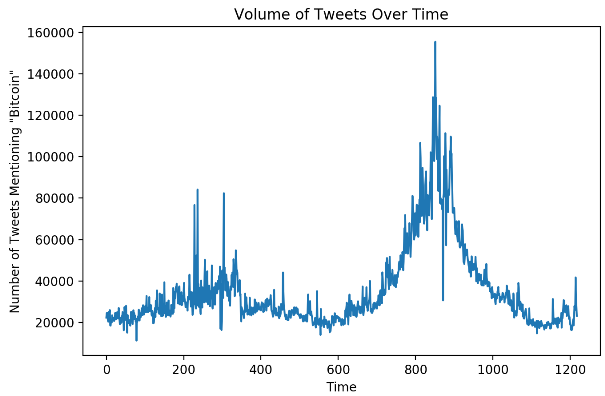

Figure 2: Volume of Tweets mentioning "Bitcoin" over time. Graph created in Python using matplotlib.

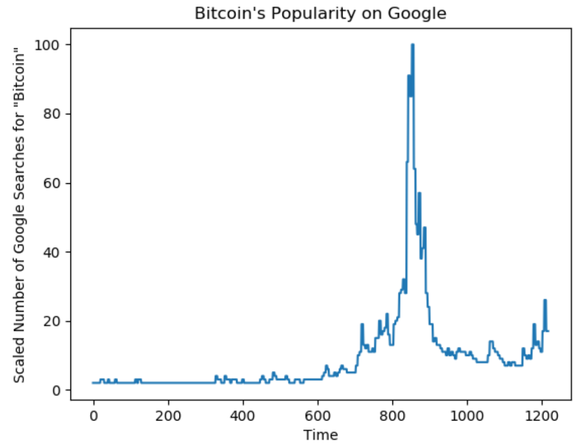

Figure 3: Scaled quantity of Google searches for "Bitcoin" over time. Graph created in Python using matplotlib.

4.2 Economic Data

Initially, we hypothesized that the economic factors would exhibit little correlation to the price of Bitcoin because of the currency's structure. Previous literature disagreed with the exact factors with the most influence on Bitcoin, causing our decision to be a challenging determination. We concluded that the most relevant economic factor was most likely a large market index. We decided to use the S&P 500, an American stock market index based on the market capitalization of 500 large companies having common stock listed on the NYSE, NASDAQ, or the Cboe BZX Exchange. Daily values were obtained from investing.com for use in the study.

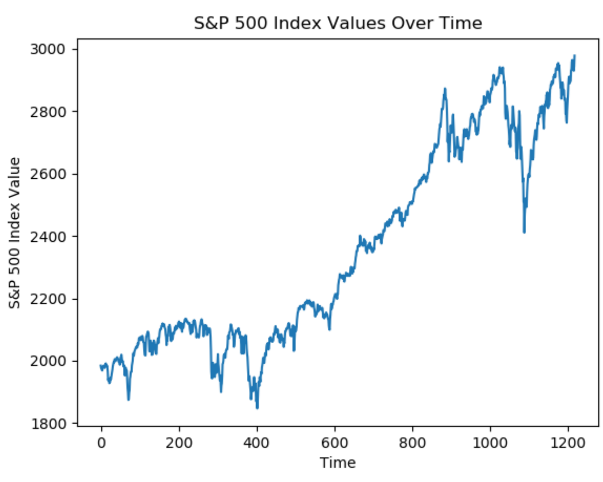

Figure 4: Value of the S&P 500 Index over time. Graph created in Python using matplotlib.

In addition to the S&P 500, we selected the exchange rates from both Euros and Japanese Yen to USD. These indicators were the next most relevant pieces of economic data because exchange rates are capable of measuring local economic growth in the context of a global economy. In other words, when examining a metric such as the GDP (Gross Domestic Product), which tracks the growth of a nation's economy, the measurement is taken on such a large scale that it is not relevant for short term studies. Exchange rates thus measure the relationship between two economies on a smaller time scale, making them more relevant to our dataset. In order to determine which currencies to compare, we considered the currencies most commonly used in Bitcoin conversions–the U.S. dollar, the Japanese yen, and the euro. Daily exchange rate data was then obtained from investing.com for Japanese Yen to USD and

Euro to USD.

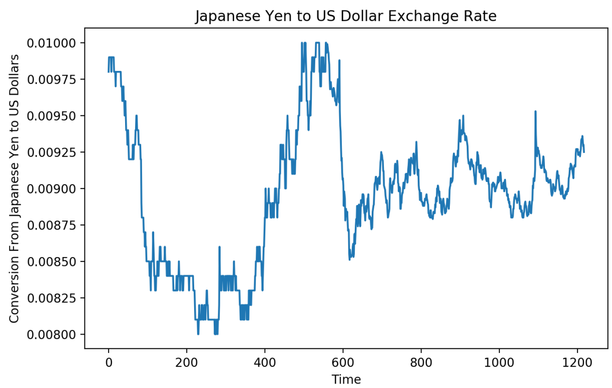

Figure 5: Japanese Yen to US Dollar exchange rate over time. Graph created in Python using matplotlib.

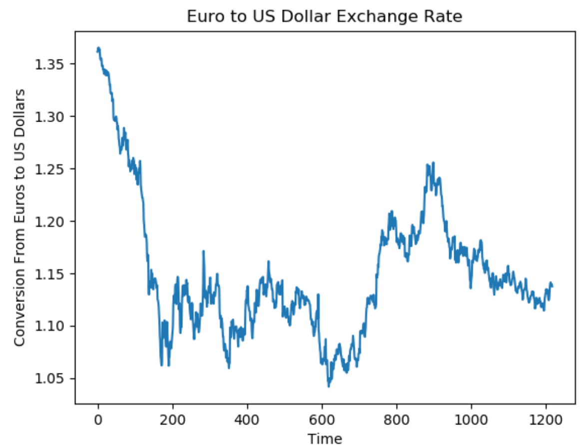

Figure 6: Euros to US Dollar exchange rate over time. Graph created in Python using matplotlib.

4.3 Data Preparation

The first aspect of the data we considered was the time scale. Since the model returns more accurate predictions with lower error when it has access to more data, we aimed to maximize the amount of training points. However, there are additional, compounding factors that can contribute to the accuracy of the model, especially when it comes to the treatment of a volatile system like Bitcoin. For example, the same feature may change its role over time, altering its effect on the actual response. From 2013 to 2016, the value of Bitcoin was relatively low and could be treated as having negligible fluctuation. Although data for the features existed for that time period, including that segment of data could unnecessarily complicate the predictive model by means of a diluted scalar. The dataset was also limited by the feature with the shortest available history. The final period was roughly July 1, 2014 to July 1, 2019. The next thing to consider while preparing the data were missing values in any of the samples. In order to fix this we wrote data parsing programs and used Microsoft Excel to sort through the data, checked for missing points, and completely removed any sample with missing data. This limited the final dataset to 1,219 samples. Lastly, we upscaled the lower resolution dataset provided by Google Trends to match the daily samples of the other features.

To ensure that our model did not overfit its data, we split our dataset into two parts: a training and a testing set. Regression was run on the training set, using three different groupings of data: one consisting of just social features (Google Trends data and Tweet volume), one on economic features (exchange rates and SP 500 Index), and one on all features. This produced three different sets of theta values. The thetas were then multiplied by the transposed test sets to produce predicted Bitcoin prices over time.

5 Results

To evaluate the error of each set of output with respect to actual data, R^2 values were computed. This statistic is commonly used to measure variance in data science, with values closer to one indicating higher accuracy while values closer to zero indicate lower accuracy. For the social, economic, and full test sets, R^2 values were 0.303, 0.489, and 0.840, respectively.

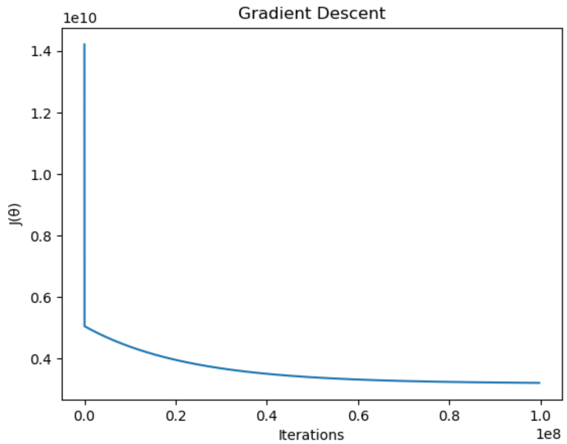

Figure 7: $J(\theta)$ plotted over iterations for training on social features. Graph created in Python using matplotlib.

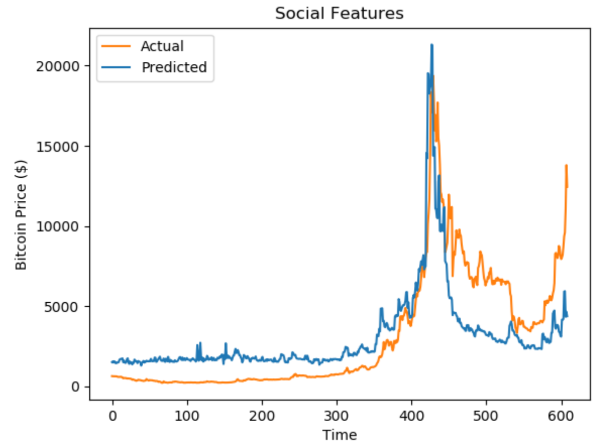

Figure 8: Predicted vs Actual Bitcoin price generated from test set of social features. $R^2 = 0.303$. Graph created in Python using matplotlib.

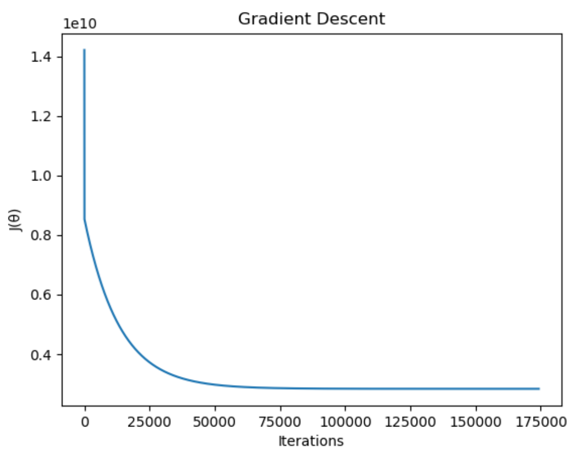

Figure 9: $J(\theta)$ plotted over iterations for training on economic features. Graph created in Python using matplotlib.

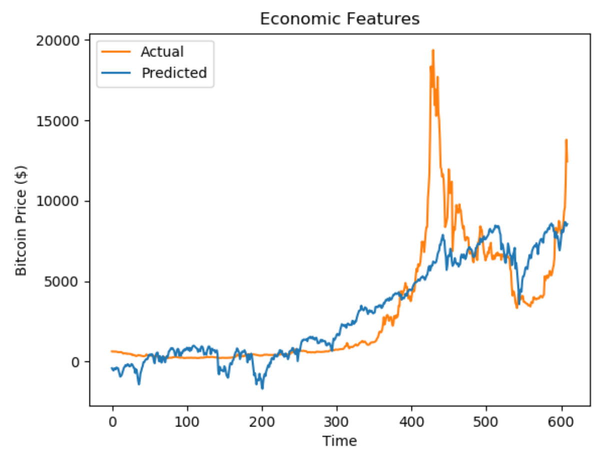

Figure 10: Predicted vs Actual Bitcoin price generated from test set of economic features. $R^2 = 0.489$. Graph created in Python using matplotlib.

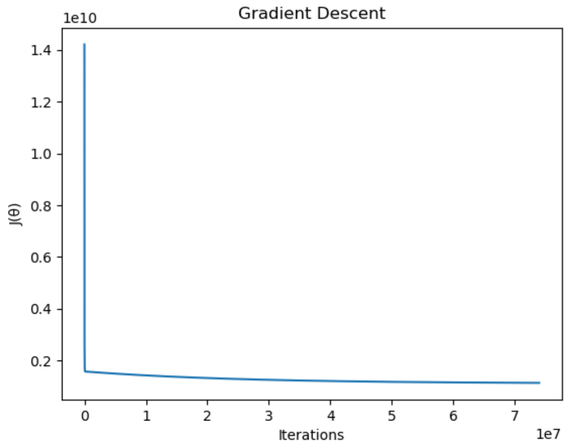

Figure 11: $J(\theta)$ plotted over iterations for training on all features. Graph created in Python using matplotlib.

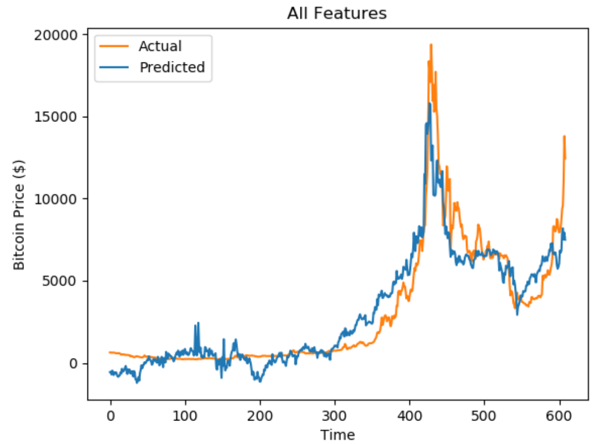

Figure 12: Predicted vs Actual Bitcoin price generated from test set of all features. $R^2 = 0.840$. Graph created in Python using matplotlib.

6 Discussion

While we initially hypothesized that social features would be the most accurate predictor of Bitcoin prices, our results indicate that this is not entirely the case. In comparison to the other two training sets, social features produced reasonably accurate depictions of short-term volatility. This is most notable between samples 300 and 450 (see Figure 8), where the prediction line appears to model the shape of the actual line moderately well. In other regions of the graph when Bitcoin prices are less volatile, such as the interval between samples 0 to 300, predicted output appears less accurate. This could indicate that social factors provide better insight about short-term rather than long-term trends. While the trend produced by the economic dataset does not appear to resemble actual Bitcoin prices as well as the social regressors (see Figure 10), its R2 value is notably higher (0.489 compared to 0.303). Therefore, economic events are better long-term predictors of Bitcoin prices, but are less reliable at predicting sudden spikes of volatility.

As expected, the use of both social and economic features provided the most accurate results, generating the highest R2 value and a trendline which most closely resembles actual Bitcoin price data (see Figure 12). While short term volatility is less accurately modeled than in Figure 8 (the all time high between samples 400 and 500 is underestimated using this dataset), the inclusion of both social and economic features acts as a more holistic approach, producing more long-term reliability.

We can verify that these results are accurate by referencing the asymptotic behavior exhibited by the cost function graphs (Figures 7, 9, and 11).

7 Conclusion

This paper provides a comprehensive analysis of the extent to which economic and social factors influence Bitcoin's value. This was done by compiling datasets which were input into a multiple linear regression algorithm that used gradient descent to determine the degree of correlation between each feature and Bitcoin's value. This study is significant for several reasons: it can be associated with marketing strategies for Bitcoin exchanges, affect investment outlook for future and current Bitcoin investors, and offers an explanation for the notoriously volatile state of Bitcoin.

Previous studies have evidenced the connection between Bitcoin's popularity on social media and its value; however, few have examined other possible factors affecting Bitcoin's price. In this paper we have shown that in fact, social factors have a lesser effect than economic indicators, though we have yet to determine the exact reason. Future research can investigate the extent to which economic indicators already account for social perception–analyzing the important economic factors convincing speculators to seek high risk investments. There are additional methods to analyze our dataset, such neural networks or polynomial regression. These may generate different results and yield alternate explanations for the data. All studies aim to minimize the risk of falsified, missing, or misleading data; therefore, training the model with different data could significantly alter the results. For instance, a future study could implement the same method but instead, use data from Reddit, Instagram, or Facebook for a more dense dataset (meaning more volume and variety). The same could be said about economic indicators. While we strategically chose the SP 500 Index and exchange rates to represent the state of our economy, other metrics, such as the Dow Jones Industrial Average or interest rates, could reveal different correlations. Future work can also generalize our results for Bitcoin to other decentralized currencies (Ethereum, Litecoin) and possibly the market as a whole through similar methods.

8 Acknowledgements

Thank you, Shadi Mohagheghi, Rachel Redberg, and Angela Zhang for teaching the mathematical concepts that were used in the paper.

9 Author Contribution Statement

J.S. implemented the experiment in python, J.H., A.L., and J.S. compiled the data, all authors conceived the experiment, collected data, and composed the paper.

10 GitHub Repository

This is a link to the GitHub repository used to conduct this study. It includes all data, code, and results: https://github.com/JackDHughes/On-the-Dynamics-of-Bitcoin-in-Relation-to-Social-and-Economic-Indicators

11 References

Abraham, Jethin, Daniel Higdon, John Nelson, & Juan Ibarra (2018) "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis", SMU Data Science Review: Vol 1: 3(1). Available at: https://scholar.smu.edu/cgi/viewcontent.cgi?article=1039&context=datasciencereview (Accessed July 21, 2019)

Aggarwal, Gourang, Kimberly Oostman, Vimal Patel, & Gaurav Varshney (2019) "Understanding Social Factors Affecting The Cryptocurrency Market", Available at: https://arxiv.org/abs/1901.06245 (Accessed July 21, 2019)

Bai, Qing, Roger H.L. Chiang, Mai Feng, Shan Zhe, & Xin Wang (2015) "The Impacts of Social Media on Bitcoin Performance", ICIS. Available at: https://pdfs.semanticscholar.org/44ab/c5e1deeb1ef1716dadc13f889a5fb45b8d69.pdf (Accessed July 21, 2019)

BitInfoCharts (2019) "Bitcoin Tweets Historical Charts", Available at: https://bitinfocharts.com/comparison/bitcoin-tweets.html (Accessed July 22, 2019)

Chang, Yale, Carl Yeung, & Chris Yip (N/A) "Analysis of the Influence of Economic Indicators on Stock Prices Using Multiple Regression", Available at: https://www.seas.upenn.edu/~ese302/Projects/Project_4.pdf (Accessed July 21, 2019)

CoinMarketCap (2019) "Historical Data For Bitcoin", Available at: https://coinmarketcap.com/currencies/bitcoin/historical-data/?start=20140701&end=20190701 (Accessed July 22, 2019)

Colaianni, Stuart, Stephanie Rosales, & Michael Signorotti (N/A) "Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis", Available at: http://cs229.stanford.edu/proj2015/029_report.pdf (Accessed July 21, 2019)

Google Trends (2019) "Daily Number of Bitcoin Search Results", Available at: https://trends.google.com/trends/explore?q=bitcoin&geo=US (Accessed July 22, 2019)

Huberman, Gur, Jacob Leshno, & Ciamac C. Moallemi (2019) "An Economic Analysis of the Bitcoin Payment System", Columbia Business School Research Paper, 17(92). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3025604 (Accessed July 21, 2019)

Investing.com (2019) "Japanese Yen to U.S. Dollar", Available at: https://www.investing.com/currencies/jpy-usd (Accessed July 22, 2019)

Investing.com (2019) "Euro to U.S. Dollar", Available at: https://www.investing.com/currencies/eur-usd (Accessed July 22, 2019)

Kjærland, Frode, Aras Khazal, Erlend A. Krogstad, Frans B. G. Nordstrøm, & Are Oust (2018) "An Analysis of Bitcoin's Price Dynamics", Journal of Risk and Financial Management, 11(63). Available at: https://www.researchgate.net/publication/328310215_An_Analysis_of_Bitcoin's_Price_Dynamics (Accessed July 21, 2019)

Lönnö, Jacob, Evita Stenqvist (2017) "Predicting Bitcoin Price Fluctuation with Twitter Sentiment Analysis", Available at: http://www.diva-portal.org/smash/get/diva2:1110776/FULLTEXT01.pdf (Accessed July 21, 2019)

Nakamoto, Satoshi (2008) "Bitcoin: A Peer-to-Peer Electronic Cash System", Available at : https://bitcoin.org/en/bitcoin-paper (Accessed July 21, 2019)

Yermack, David (2013) "Is Bitcoin a Real Currency? An Economic Appraisal", NBER Working Paper, 19747. Available at: https://www.nber.org/papers/w19747 (Accessed July 21, 2019)