Statistical Analysis of Cryptocurrencies

Borude Nishant, Chakradeo Mihir, Patil Dhanashri, Sonawane Bhushan, Yele Aditya

I. Introduction

With the advent of blockchain technology, cryptocurrencies have gained a lot of importance in today's digital world. As cryptocurrencies have direct and indirect effects on multiple domains, we present statistical analysis to validate the same. These effects being bi-directional and volatile, the analysis becomes intriguing as well as challenging.

We started off with parametric inference of the cryptocurrency data set to get an idea of the distribution. The data being sequential in nature, we decided to perform Time Series Analysis for the prices of different cryptocurrencies.

Furthermore, to understand what factors affect cryptocurrency prices, we studied how Google Searches affected Cryptocurrency prices using Linear Regression, Wald's test and KS test. We also applied Multiple Linear Regression to find out which feature of cryptocurrencies was most influenced by the Google Searches.

Any currency has some impact on the Finance sector, so we decided to put cryptocurrencies to that test. Specifically, we studied how companies producing hardware for mining cryptocurrencies performed with bitcoin prices. We applied Multiple Linear Regression for predicting the stock values of these companies. Furthermore, we came across articles which claimed that the number of bitcoin investors in Greece increased after the crisis of 2015[1]. So, we studied this by applying Simple Linear Regression.

Ransomware is a cyber attack where hackers encrypt files on a system and demand money for unlocking them. In recent years there has been an increase in the number of ransomware attacks. We attribute this increase to the fact that cryptocurrencies are untraceable. Hence, we hypothesize that the increase in number of bitcoin transactions is because of this increase in ransomware attacks. We applied Wald's Test for analyzing the same. Further we did parametric inferencing to find the parameters of the distribution. Given the normal distribution for prior and likelihood, we applied Bayesian Inference to find the posterior distribution of the data subject to new random data generated from the same distribution.

Expenditure in Gambling has been increasing with increase in total bitcoins. News-bitcoin's survey[2] states that more than $4.5 billion have been wagered in bitcoin since 2014. Bitcoin transaction being untraceable and secured is likely to motivate gamblers to move to online gambling platforms such as crypto-games[3]. Hence, we studied how much players are spending in different gambling games such as Wagering, EGM, Casino, Keno and Lottery. We applied Wald's test, Permutation test and Simple Linear Regression.

II. Dataset

A. Cryptocurrency Dataset

The Cryptocurrency dataset [4] from Kaggle is our main dataset, which consists of daily information about the price, market capital, mining difficulty and rest important features of bitcoin over the period of 2010 to 2018. The main table consists of 2920 rows and 24 feature columns. Similar datasets are present for other 12 currencies viz. ethereum, ripple and neo.

Date	btc market price ($)	btc total bitcoins	btc n transactions
2018-02-17	10841.991	16871012.5	173789
2018-02-18	10503.298	16873037.5	159495
2018-02-19	11110.965	16875062.5	187367
2018-02-20	11390.391	16876825.0	198455

TABLE I
SNAPSHOT OF CRYPTOCURRENCY DATASET

B. Google Trends Dataset

We got the Google Search data from Google Trends [5]. This data has the date and the number of searches of the keyword "Bitcoin" for that corresponding date. Here, the search count is an adjusted number, which is relative to the maximum number of searches, where 100 represents the maximum searches for that date range. Total number of rows: 261, which consists of data from 2013 to 2018. Refer Table II for snapshot of dataset.

Date	Searches
2013-04-28	3
2018-02-18	26

TABLE II
SNAPSHOT OF GOOGLE TRENDS DATASET

C. Finance Dataset

The stock dataset[6] consists of daily stock prices of three companies NVIDIA(1999-01-22 - 2017-11-10; 4734 rows), AMD(1983-03-21 - 2017-11-10; 8738 rows), Taiwan Semiconductor Manufacturing(2005-02-25 - 2017-11-10; 3202 rows). Each row consists of the following columns date, opening price, closing price, high, low and the volume of shares for that day. The description of each column is given below.
(1) Date: The day the specific stock data refers to
(2) Open: Opening price of the stock on that day
(3) High: Highest price of the stock on that day
(4) Low: Lowest price of the stock on that day
(5) Close: Closing price of the stock on that day
(6) Volume: Number of stocks traded on that day

Date	Open	High	Low	Close	Volume
17/11/09	205.27	206.33	200.37	205.32	23895006
17/11/10	213.08	218.67	211.63	216.14	31300857

TABLE III
SNAPSHOT OF STOCKS DATASET FOR NVIDIA STOCKS

The other dataset used was the GDP of Greece[7]. It contains the year quarter and the increase or decrease in GDP for that quarter in percentage. The dataset starts from the first quarter of 2013 to the first quarter of 2017 and consists of 18 rows.

Time	Value
2013-Q1	-2.192256
2013-Q2	0.077293

TABLE IV
SNAPSHOT OF GREECE GDP

D. Ransomware Dataset

We scraped the Ransomware dataset[8] from Symantec website. The data consists 31 rows from Jan 2015 to July 2017 that states the different number of attacks that took place in a particular month. Table V shows the format of this dataset.

Month Year	Type of Attack	Ransom Amt	Attack
May 2016	Crypto	2 BTC	Mischa
May 2016	Crypto	5 BTC	Bucbi
May 2016	Crypto	0.42 BTC	Enigma
April 2016	Crypto	0.24 BTC	Rokku
March 2016	Crypto	$300	Cryptohasyou

TABLE V
SNAPSHOT OF RANSOMWARE DATASET

The first column is the date. Second is the type of attack whether it's Crypto or Locker. The third column states the ransom amount in USD/BTC/Yen etc. And the last column is which attack took place that time.

E. Gambling Dataset

Gambling dataset[9] describes monthly player expenditure in following gambling streams- Wagering, EGM, Casino, Keno and Lottery. Dataset consists of three fields- Moth Year, Game Stream and Player expenditure. Refer Table VI for snapshot of dataset.

Month Year	Game Stream	Player expenditure($)
July 2004	Casino	45662132.51
July 2004	EGM	145766780.77
July 2004	Keno	6836926.58
July 2004	Lottery	32451660.78
July 2004	Wagering	27298552.41

TABLE VI
SNAPSHOT OF GAMBLING DATASET

Dataset is collected from Queensland Government website. Following are the definitions of the dataset provided by the Office of Liquor and Gaming Regulation:
(1) Month Year: The month and year from which the gambling data is provided.
(2) Game Stream: The gambling type, which may be Casino (including electronic gaming machines and Keno in casino venues), EGM (electronic gaming machines in clubs and hotels), Keno (in clubs, hotel and other non-Casino venues), Lottery or Wagering.
(3) Player expenditure: The amount of money lost by players. Also referred to as metered win, commission or revenue in different parts of the gambling industry.

III. DATA PRE-PROCESSING

A. Cryptocurrency Dataset

1) The Date from cryptocurrency data was in String format, so we converted it to datetime format
2) Merge Join: The number of rows in tables of different currencies were not equal. In order to find correlation between different currencies we performed merge join on the date column

B. Google Trends Dataset

1) The Date from Google Trends Data was in String format, so we converted it to datetime format
2) Scaling: The bitcoin price data was in a very different range (100-11000) compared to the search data (1-100). So, we scaled the bitcoin price in the range of 0-100.
3) Moreover, for showing asymptotically normal (for Wald's test), we took average over the months. For that, we split the Date feature column over Month and Year. Refer Table VII for snapshot of preprocessed dataset.

Searches	btc_m_price	Year	Month
3	0.701172	2013	4
26	55.127267	2018	2

TABLE VII
SNAPSHOT OF PREPROCESSED GOOGLE TRENDS DATASET

C. Stock Dataset

1) For using this dataset we had to extract only the data for the timeline for which corresponding bitcoin data was available.
2) Normalization: We normalized the data in the btc_market_price, btc_total_bitcoin, btc_trade_volume, btc_n_transaction columns of the bitcoin dataset rescaling the values between 0 and max-min.
3) Precalculation: Since the data for GDP of Greece is available for every quarter and the data for bitcoin marketprice is available daily we had to average over the market price for every quarter.

D. Ransomware Dataset

1) The Date from Ransomware Data was in String format, so we converted it to datetime format
2) Averaged the number of attacks that took place in a particular month
3) Dropped the ransom amount column
4) Split the date into Month and Year
5) Normalization: Furthermore, when comparing this dataset for any further analysis, we normalized it's count entries to the scale [0,1] in order to make it comparable with the normalized bitcoin dataset.

Year	Month	Count
2016	4	14
2016	5	13
2016	6	13
2016	7	5
2016	8	9

TABLE VIII
SNAPSHOT OF PROCESSED RANSOMWARE DATASET

E. Gambling Dataset

1) We split Month Year from one column to individual separate columns.
2) There are five gambling streams in dataset. Every month has player expenditure for respective month. We modified our dataset making every gambling stream separate column.
3) Normalization: We also normalized the dataset to 0-1 scale for comparing with bitcoin dataset.

IV. HYPOTHESES

A. BITCOIN PRICE ANALYSIS AND EFFECT ON OTHER CURRENCIES

We studied our primary data set - the Cryptocurrency data set, for getting a general idea of the distribution and thereafter applied techniques learned in class.

*1) **Estimate the mean of bitcoin price using MME:*** By plotting the graph of CDF of bitcoin closing price we observed that it looked like an exponential distribution. So, based on this educated guess, we applied MME, and calculated the parameter λ:

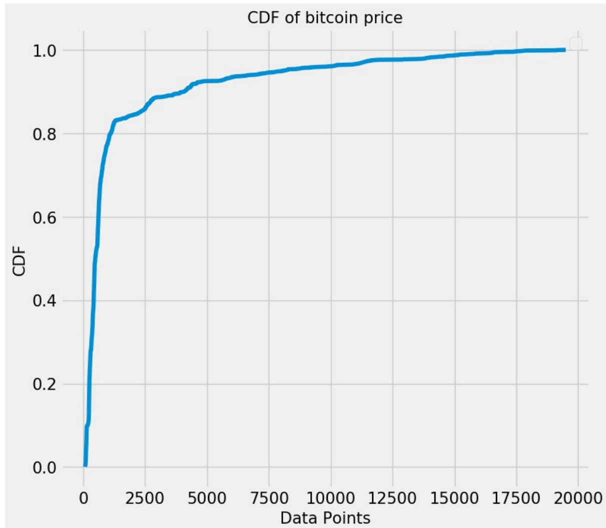

Fig. 1. CDF of Bitcoin price

i. No. of unknown parameter(s) k = 1

ii.
$$\hat{\alpha}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1}$$

iii.
$$E\big[exponential(X)\big] = \frac{1}{\lambda} \tag{2}$$

iv. Equating the above moments we get,

$$\frac{1}{\hat{\lambda}_{MME}} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{3}$$

$$\hat{\lambda}_{MME} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i} \tag{4}$$

$$\hat{\lambda}_{MME} = 0.00067 \tag{5}$$

*2) **Time Series Analysis of Bitcoin closing price:*** We split the training and testing data in the ratio 4:1. We applied following three techniques:

i. **EWMA**

We tried different values of α for EWMA model, wherein α = 0.8 gave the most accurate prediction. The graph for EWMA plot is in Figure 2

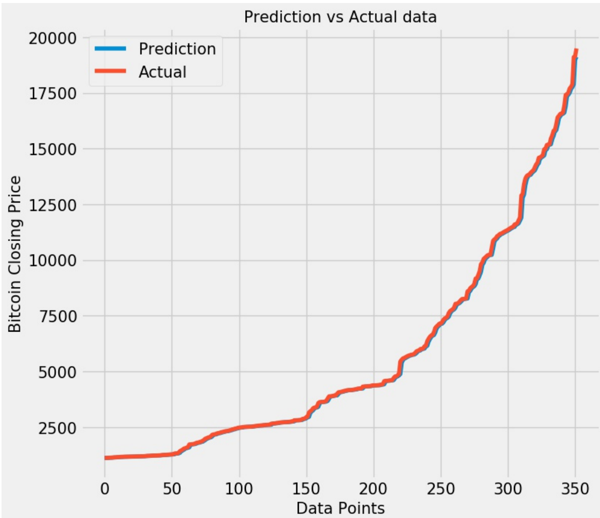

Fig. 2. Time Series Analysis using EWMA

ii. **Auto Regression**
Given that the data is sequential we can apply AR model. We tried different values of α, we received less MAPE for α = 300 as seen in Table IX. The graph can be seen in 3

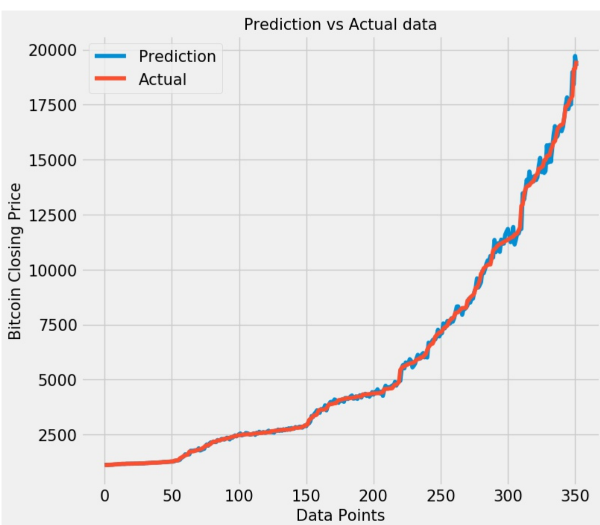

Fig. 3. Time Series Analysis using AR model

iii. **Seasonal**
Here also we tried different values of α(Table X), we received less MAPE for α = 100. Figure 4 shows the graph.

Conclusion: Table XI shows the results. EWMA gives a better estimate for the bitcoin closing price since the fluctuation in data values is less and by keeping α value large we are giving more weight to immediate previous actual values

Alpha value	MAPE
200	1.4756
300	1.4582
400	1.5195

TABLE IX
VARIATION IN MAPE WITH DIFFERENT VALUES OF ALPHA FOR AR

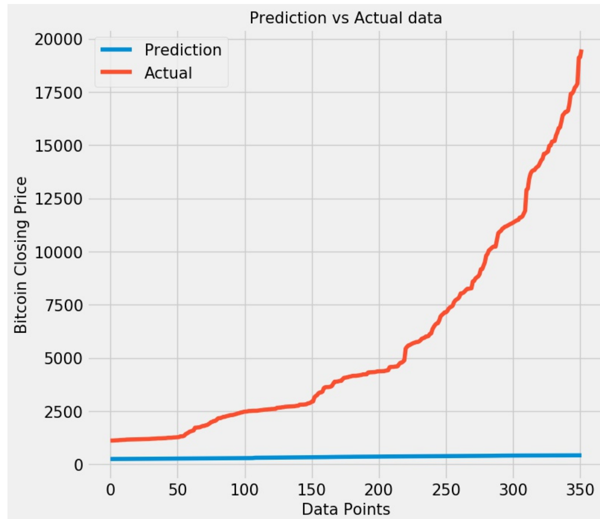

Fig. 4. Time Series Analysis using Seasonal

3) ***Predict ethereum closing price based on bitcoin dataset:*** After plotting correlation matrix between ethereum closing value and all the features of bitcoin dataset we observed that ethereum closing value is strongly correlated with below three features of bitcoin dataset:
X1, X2, X3 vectors : miners_revenue, hashing_difficulty, bitcoin_market_price
Y vector : ethereum_closing_value
By applying Multiple Linear Regression, we get, SSE and MAPE values as shown in Table XII

Conclusion: Even though given three bitcoin dataset features are closely correlated with ethereum closing value, they do not give good prediction for ethereum closing value

B. EFFECT OF GOOGLE SEARCH ON BITCOIN PRICE

We plotted the Frequency of Google Searches of the keyword "Bitcoin" from 2014-2018, and also plotted the Bitcoin market price for the same years 2014-2018. Surprisingly, we observed that those plots almost coincided, as seen in figure 5

Alpha value	MAPE
100	79.7038
500	79.8672
1000	89.2542

TABLE X

VARIATION IN MAPE WITH DIFFERENT VALUES OF ALPHA FOR SEASONAL

Method	MAPE
EWMA	0.9990
AR	1.4582
Seasonal	89.6630

TABLE XI

TIME SERIES ANALYSIS RESULTS

*1) **Bitcoin market price and Frequency of Google Searches follow the same distribution**:* Here, we prove that the Bitcoin Market Price and Frequency of Google Searches follow the same distribution. We used following techniques: KS test and Wald's test.

- **KS Test:** Hypothesis: H_0: Both Bitcoin market price and Google Search Frequency come from the same distribution.
  We plotted the CDF of both Bitcoin market prices and Google Search Frequency, which can be seen in figure 6.
  Statistic: D = 0.117 for $N = 50$ data points. So, for $\alpha = 0.05$, we have a C value of $C = 0.188$ (obtained from the table from real-statistics website [10]). As $D > C$, we can say that we accept the Null Hypothesis.
  Conclusion: Both, Bitcoin Market Price and Frequency of Google Searches follow the same distribution.

- **Wald's Test for $\alpha = 0.05$:**
  Hypothesis: H_0: Both Bitcoin market price and Google Search Frequency come from the same distribution.
  Assumptions: For Wald's test, we need the data to be Asymptotically Normal. We consider the average number of searches over months. Also, we averaged out the bitcoin market values over months. Thus, using CLT, we say that both the datasets used for this analysis are asymptotically normal.
  The results are in Table XIII
  For a 2 tailed Wald's test, with $\alpha = 0.05$, we have $Z_{\frac{\alpha}{2}} = 1.96$
  Conclusion: We can see for alpha = 0.05, for this particular test, the W value is less than $z_{\frac{\alpha}{2}}$. Hence we accept the Null hypothesis.

*2) **Predicting Bitcoin Market Price from Google Searches**:* Having established that the two plots coincide

SSE	MAPE
1831637.40	777.97

TABLE XII

MULTIPLE LINEAR REGRESSION RESULTS

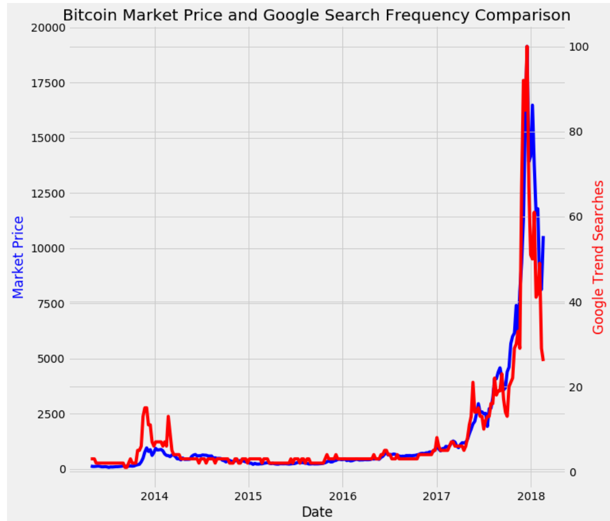

Fig. 5. Google Trends Searches and Bitcoin Market Price from 2014-2018

and come from the same distribution, we decided to test if we can create a Linear Regression model to fit the data, so that given the number of searches in a month, we can give the estimated Bitcoin market price. Here, Y = Bitcoin Price, X = Frequency of Google Searches. The plot of the Linear Regression fit can be seen in 7 We also calculated the SSE and MAPE metrics:
SSE: 2051.36, **MAPE:** 52.93
Conclusion: The MAPE and SSE are very large, which means that we cannot actually fit a Simple Linear Regression Model based on only one feature (frequency of Google Search keywords).

*3) **Which Bitcoin feature is most influenced by Google Searches**:* We tested out three features, namely, (1) bitcoin number of transactions, (2) bitcoin market price and (3) bitcoin trade volume. We applied Linear Regression three times, where, Y = one of the three features, and X = Frequency of Google Searches. We observed the following beta values:
Conclusion: As seen in table XIV, It is evident from the result that the number of transactions is influenced the most by Google Searches

C. BITCOIN AND FINANCE

*1) **Predicting Stocks using cryptocurrency data set**:* Cryptocurrency mining requires specialized hardware Application Specific Integrated circuits (ASIC) and Graphical Processing Units (GPUs). This means if there

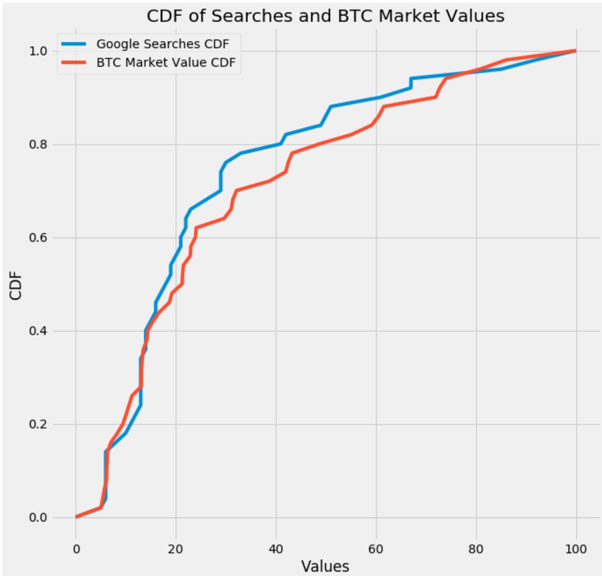

Fig. 6. CDF of Searches and CDF of Bitcoin market price vs values

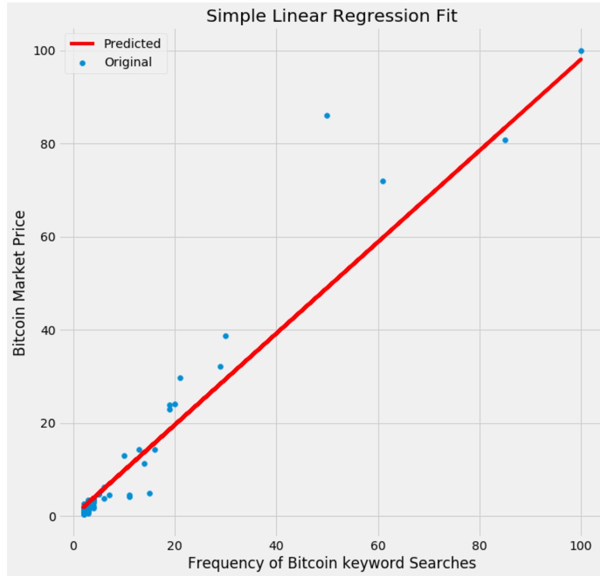

Fig. 7. Simple Linear Regression Fit

W value	p value
0.2812	0.7785

TABLE XIII
WALD'S TEST RESULTS

Transactions	Price	Trade_Vol
1.643	1.060	0.758

TABLE XIV
BETA VALUES

is any change in the metrics of bitcoin there will be change in the stocks of ASIC and GPU manufacturing companies. We propose that there is a linear relationship between stock prices of these companies and certain metrics of bitcoin. Using these metrics we predict the value of stocks of three ASIC and GPU making companies-NVIDIA, AMD, Taiwan Semiconductor Manufacturing (TSM) and try to find the factor which most affects the stock price.
Method: Multiple Linear Regression
Features: btc_market_price, btc_total_bitcoin, btc_trade_volume, btc_n_transaction
Prediction: Stock price of an ASIC and GPU making company
Results: The SSE and MAPE for all three companies are given below in table XV. The Yhat vs Y actual graph for TSM is given in figure 8. In table XVI we can see the weights of the features for TSM which best fits the data. The feature btc_market_price has the highest weight indicating that the value of the stock is more dependent on the market price as compared to others. For Nvidia and AMD as well market price has the highest weights **1.1611** and **1.3941** respectively.

Conclusion: From the graph in 8 we see that stock price of TSM most fits the data with least values of MAPE and SSE
Bitcoin features linearly fit stock prices with highest dependency on bitcoin market price

2) ***Similarity between stock price distribution of TSM and Bitcoin:*** In the previous hypothesis we saw that the TSM and bit coin data are linearly related. We extend this further to see how similar are the distributions of stock prices of TSM and market price of bitcoin.
Method: Permutation Test
Conclusion: We applied permutation test and found that for all the permutations the mean of the observations was less than T_{obs} and so the p-value was **0**, which is less than α = 0.05. Therefore the distribution of market price and stock prices of TSM is not same.

3) ***Effect of Financial Crisis on Bitcoin price:*** Here we take the instance of Greece and try to see if there is any relation between the recent financial crisis in Greece to the price of bitcoin. We try to see with decreasing belief in their own currency did people in Greece start investing in bitcoin. For checking our hypothesis we predicted the market price of bitcoin.
Method: Simple Linear Regression
Features: GDP of Greece
Prediction: Market price of bitcoin
Conclusion: As we can see from figure 9, bitcoin data does not fit linearly with the GDP of Greece and therefore we conclude that there is no relationship or the decrease in GDP has very little effect on the increase in

Company	SSE	MAPE
NVIDIA	1.4475	126.0607
AMD	9.2806	103.6849
TSM	1.090	23.0741

TABLE XV
SSE AND MAP VALUES OF THE MULTIPLE LINEAR REGRESSION TO PREDICT STOCK PRICES OF THREE COMPANIES NVIDIA, AMD AND TSM

Company	Weights
btc_market_price	0.6518
btc_total_bitcoins	0.2934
btc_trade_volume	-0.1899
btc_n_transactions	0.3536

TABLE XVI
SSE AND MAP VALUES OF THE MULTIPLE LINEAR REGRESSION TO PREDICT STOCK PRICES OF THREE COMPANIES NVIDIA, AMD AND TSM

bitcoin market price. The values of SSE and MAP are **1.1239**, **175.8962** respectively.

D. *CRYPTOCURRENCY AND RANSOMWARE ATTACKS*

*1) **The number of ransomware attacks and the bitcoin cost per transaction follow the same distribution**:* We already discussed how ransomware attacks could've affected the cryptocurrency market. Based on this, we hypothesize that the bitcoin dataset and ransomware data could be directly related. We considered two fields from each data and proposed that the number of attacks from Jan 2015 to July 2017 and the bitcoin transaction cost are from the same distribution.

H_0: Bitcoin transaction cost and number of ransomware attacks are from the same distribution.

Method: Wald's test for alpha = 0.05 **Assumptions:** Data should be asymptotically normal. We consider the average number of attacks that take place. Also, we averaged out the the entries in the bitcoin transaction dataset. Thus, using CLT, we say that both the datasets used for this analysis are asymptotically normal.

The results are in Table XVII

W value	p value	C.I.
0.5581	0.5767	[-0.0796, 0.1430]

TABLE XVII
WALD'S TEST RESULTS

Conclusion: We can see for α = 0.05, for this particular test, the w value is less than $z_{\frac{\alpha}{2}}$. Hence we accept the Null hypothesis.

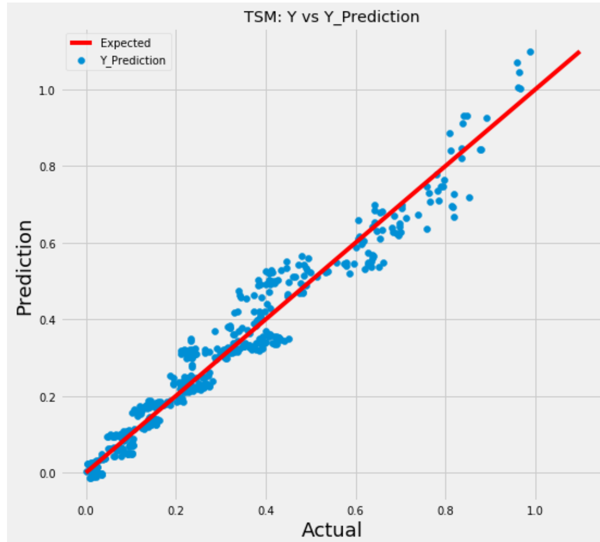

Fig. 8. YHat vs Y for TSM

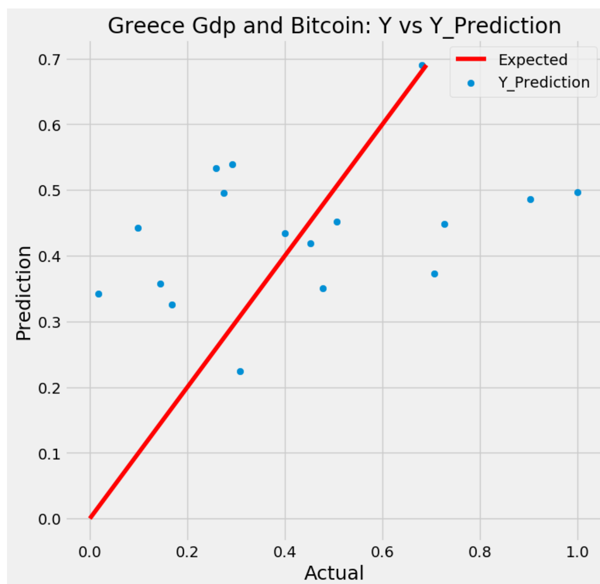

Fig. 9. YHat vs Y for Greece and Bitcoin regression

*2) **Parametric Inference for normally distributed data**:* We proved using CLT that the data is asymptotically normal. On plotting the graph for the number of ransomware data it seems to follow the bell curve as shown in Figure 10.

We did parametric inferencing for this distribution to find μ_{MLE} and σ_{MLE}

Method: MLE
We use MLE to find the parameters of the Gaussian model. We find the loglikelihood of for the data and apply MLE. Solving that, we get the following MLE parameters for the model:

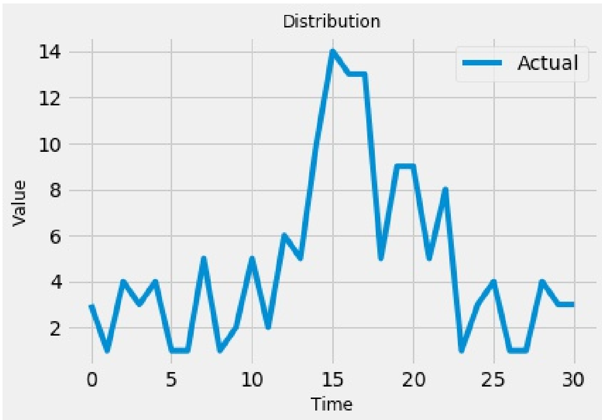

Fig. 10. Number of attacks distribution

$$\mu_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}, \sigma^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_{MLE})^2 \quad (6)$$

Conclusion: The derived parameters are

$$\mu_{MLE} = 0.3433, \sigma_{MLE} = 0.2656 \quad (7)$$

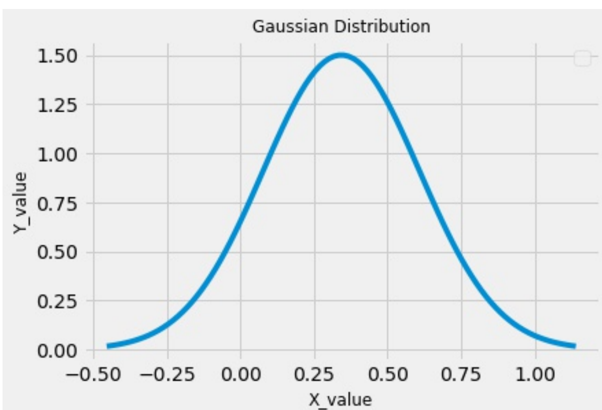

Fig. 11. Gaussian Distribution for MLE parameters

*3) **Posterior distribution of the number of ransomware attacks is normal:*** **Method:** Bayesian Inference.
Based on the previous hypothesis, we move a step further to find the posterior distribution of the data when new data is introduced randomly from the same normal distribution.

Observation: Based on what we learned in class, if the prior and likelihood is Normal, we see that the posterior distribution is also Normally distributed with updated parameters as shown in Figure 12

Varying for different values of sigma for the update, we observe that when sigma is large, then the posterior distribution doesn't change by much. But when it is low, we see the posterior distribution changing.

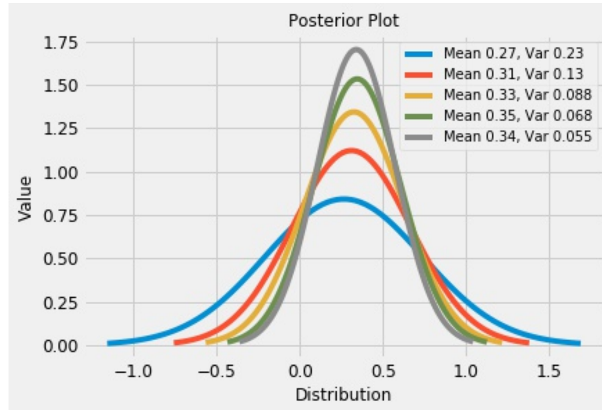

Fig. 12. Posterior Distribution for New Data

E. CRYPTOCURRENCY AND GAMBLING

*1) **Use of bitcoin have steadly risen with increase in gambling expediture:*** As discussed in introduction, use of bitcoin in gambling is increasing. For verifying the same we test following hypothesis.

H_0: Number of bicoins per year are increasing with player expenditure in various gambling streams.

Method 1: Wald's test for alpha = 0.05 **Assumptions:** Data should be asymptotially normal. We consider the average number bitcoin per month over year. We also averaged out the the entries in the bitcoin transaction dataset. Thus, using CLT, we can say that both the datasets used for this analysis are asymptotically normal. Figure 13 shows CDF of Casino.

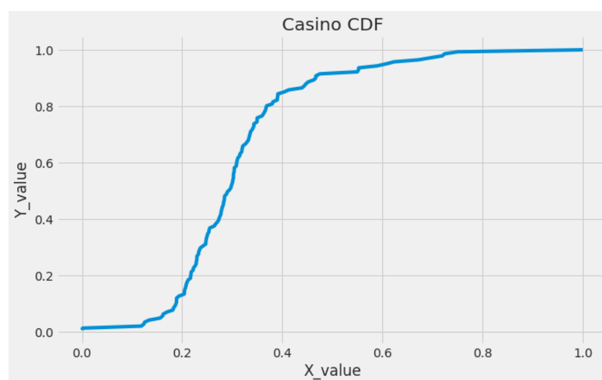

Fig. 13. CDF of Casino: Monthly averaged over year

Method 2: Permutation test We also applied Permutation test for verifying hypothesis for each gambling stream.

Results of Wald's and Permutation test are noted in table XVIII

*2) **Expenditure in Gambling have contributed into increase in price of bitcoin:*** Our first hypothesis validates that increase in gambling expenditure is indeed

Gambling Stream	Wald's Test (P-Value)	Confidence Interval	Perm Test (P-Value)	Conclusion
EGM	0.175	[-0.027, 0.151]	0.1892	Accepted
Lottery	0.0165	[0.019, 0.194]	0.0185	Rejected
Keno	0.0063	[0.037, 0.229]	0.0082	Rejected
Wagering	0.4034	[0.049, 0.121]	0.4111	Accepted

TABLE XVIII
RESULTS OF HYPOTHESIS 1: NUMBER OF BITCOINS AND EXPENDITURE IN GAMBLING

contributing to increase in number of bitcoin per year. But, how does it also contribute to increasing bitcoin prices? To answer this question, we propose following hypothesis:

H_0: Bitcoin price has been rising with increase in player expenditure in gambling.

We can prove asymptotially normality as proved in first hypothesis. We applied Wald's and Permutation test for validating the same. Table XIX shows the results.

Gambling Stream	Wald's Test (P-Value)	Confidence Interval	Perm Test (P-Value)	Conclusion
Casino	0.5177	[-0.058, 0.116]	0.5253	Accepted
EGM	0.0163	[0.021, 0.209]	0.01962	Rejected
Lottery	0.6303	[0.074, 0.123]	0.6366	Accepted
Keno	0.0010	[0.067, 0.268]	0.0017	Rejected
Wagering	0.6644	[-0.067, 0.106]	0.6705	Accepted

TABLE XIX
RESULTS OF HYPOTHESIS 2: BITCOIN PRICE AND EXPENDITURE IN GAMBLING

*3) **Predicting total number of bitcoins from Expenditure in Casino**:* From hypothesis 1, we know that monthly expenditure in Casino and number of bitcoins per month are from same distribution. Hence, we tried to fit Simple Linear regression model to fit the data, so that given expenditure in particular month, we can estimate total bitcoins added into system in that particular month. Figure 14 shows regression fit on test data.

SSE and MAPE for regression analysis are presented in table XX

SSE	MAPE
0.1775	27.5995

TABLE XX
SIMPLE LINEAR REGRESSION FOR PREDICTING TOTAL NUMBER OF BITCOINS

V. PRIOR WORK

People have previously worked on different domains related to Cryptocurrency analysis. We found the following previous implementations:

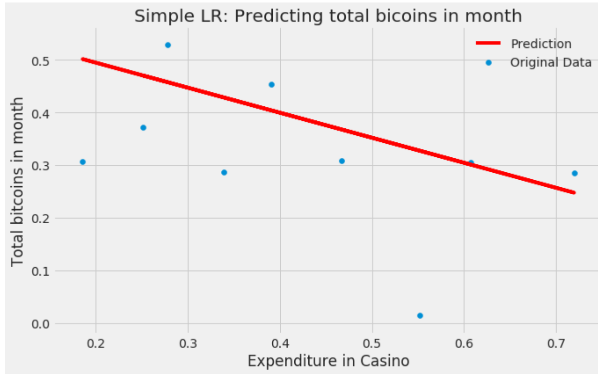

Fig. 14. Regression fit for total bitcoins with expenditure in Casino

A. Time Series Basic Analysis

The Time Series Basic Analysis [11] focuses on the bitcoin price file from the data. As opposed to our approach, the author has applied ARIMA, Dickey Fuller Test, decomposition of time series and autocorrelation. The approach is to plot different data i.e. monthly or daily data to gain insights and then apply time series analysis.

The author plotted Seasonality graph and worked out Dickey Fuller Test on the data to find the conclusion that the daily data is too stacked for finding seasonality pattern compared to monthly data. On applying DF test, they conclude that the time series is Stationary and can be used to apply forecasting techniques such as ARIMA, EWMA.

Another approach on Time Series Analysis by other author is pretty basic which shows by plotting different closing prices, which one has the highest closing price. Our approach does some extensive analysis using EWMA, AR, Seasonal model to show which model fits perfectly for this data.

B. Crypto-Correlation

Many people have studied the correlation of cryptocurrencies. For example, DanBarkhorn's work [12] where, he finds cross correlation between two time-series using a lag. And, for testing stationarity, applied unit root test (Augmented Dickey-Fuller Test).

Our work is not restricted to finding correlation between cryptocurrencies. We also applied parametric, non-parametric inferencing techniques, as well as Multiple Linear Regression for predicting Ethereum prices using Bitcoin features.

C. Analysis of Bitcoin Prices and Google Trend Searches

The exploratory analysis of bitcoins [13] have performed some exploratory analysis of Bitcoin prices and it's correlation with Frequency of Google Searches of keywords relating to Bitcoins.

The main difference in their analysis and ours is that they did correlation analysis as their basis for other hypotheses. Whereas, we have compared their distributions, and built up upon that.

Both of our analysis includes Regression Analysis, but the difference lies in the Error metrics. Their analysis talks about the R squared and P values, whereas, our analysis uses SSE and MAPE

D. Cryptocurrency and Financial Data Analysis

The paper[14] shows work on comparing different cryptocurrencies with the US Dollar. The authors considered the log likelihood of the data and applied the t test for comparison. We applied the permutation test for comparing the cryptocurrency price with stock market data of ASIC producing companies. In the paper, they conclude that none of the distributions gave best jointly fit, which is similar to our analysis.

VI. Future Work

A. Relation between National Currency and bitcoin

We analyzed and tried to establish a relation between decrease in GDP of Greece(a country in financial crisis) and increase in investment of bitcoin. Here however our predictions were constrained by the availability of data(we only had data annual GDP data of Greece from 2010 to 2016). Overtime as more data becomes available we would be able to make more accurate predictions. We can also consider other indicators such as quarterly debt to GDP ratio and try to predict an increase or decrease in bitcoin prices. This model can also be applied to other countries or to predict a financial crisis in a country should the bitcoin investment in that country spike.

B. Using other social media platforms for sentiment analysis

We studied the effect of Google searches on bitcoin. As a next step we would like to do the same study on using data from other social media platforms like Facebook and Twitter. This would also enable us to see user sentiment on which platform has the biggest effect on the bitcoin market. We would also like to do the same study for other cryptocurrencies.

VII. Source Code

Our implementation can be found here:
https://github.com/bhushan23/Cryptocurrency-Analysis

References

[1] Greeks are rushing towards bitcoins. *http://money.cnn.com/2015/06/29/technology/greece-bitcoin/index.html*.

[2] Bitcoin Gamblers Have Wagered $4.5 Billion in BTC Since 2014. *https://news.bitcoin.com/bitcoin-gamblers-wagered-4-5-billion-btc-2014/*.

[3] Crypto-Game Website. *https://www.crypto-games.net/*.

[4] Cryptocurrencies dataset. *https://www.kaggle.com/kardesh/bitcoin-blockchain-analysis/data*.

[5] Google Trends Frequency of Bitcoin Keyword Searches. *https://trends.google.com/trends/explore?q=Bitcoin/*.

[6] Stock-Data from Kaggle. *https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs/*.

[7] Quarterly GDP of Greece. *https://data.oecd.org/gdp/quarterly-gdp.htm*.

[8] Ransomware dataset. *http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/ISTR2016_Ransomware_and_Businesses.pdf*.

[9] Gambling-Dataset Queensland Government. *https://data.qld.gov.au/dataset/all-gambling-data-queensland*.

[10] Kolmogorov-Smirnov Table. *http://www.real-statistics.com/statistics-tables/kolmogorov-smirnov-table/*.

[11] Time Series Analysis of Bitcoin Data. *https://www.kaggle.com/sandipdatta/time-series-basic-analysis*.

[12] Prior work crypto-correlation. *https://www.kaggle.com/dbarkhorn/crypto-correlation*.

[13] Can we predict bitcoin price with google trend? *https://blog.exploratory.io/can-we-predict-bitcoin-price-with-google-trend-7ae111ef1f85*.

[14] A statistical analysis of cryptocurrencies. *https://www.mdpi.com/1911-8074/10/2/12/pdf*.