**Cryptocurrency Demystified**

Group C1G4

Deepika Jindal (Graduate Student at Purdue University)

1. **Introduction**

* Research Topic

In this research, we want to find out what are the dominant factors that affect the market capital of cryptocurrency market. Based on the research, we hope to find out whether the cryptocurrency market is really a market for opportunists and the other market factors that influence the cryptocurrency’s performance.

* Why Interesting

The reason that we chose this topic is that, during last year, investors in crypto market made great fortune. We want to test what are the factors that would influence this market, and what should be our strategies to invest safely.

* Background

The cryptocurrency has been a popular topic in the investment market recently and Bitcoin is one of the most successful in the market. However, after Bitcoin hit the turbulence this year, there are suspicions about whether the cryptocurrency market is just the market for opportunists.

Numerous studies on predictions of cryptocurrencies’ prices were carried out. A prediction model was conducted using Bayesian Neural Network (BNN) to predict Bitcoin price with various parameters including stock indexes, exchange rate, and hash rates [1]. Taking into consideration of 16 parameters, the BNN method is a good method predicting log price of Bitcoin.

* What we did:
* A comparative analysis between the past market and the current market to see the status quo of the market
* Regression analysis to find out what factors could influence cryptocurrencies’ market cap
* Comparison analysis of the relationship between BTC and top currencies / BTC and newly introduced currencies
* Clustering the currencies using K-means basis the coefficients of different factors obtained in regression.
* A simulation prediction to predict BTC’s future price
* Variables tested:

|  |  |
| --- | --- |
| Variables | Implications |
| Daily Opening Price of each currency | whether high opening price affects market capitalization |
| Daily Trading Volume of each currency | Whether high trading volume affect market capitalization |
| Daily Standard and Poor (S&P) 500 Index | Whether any correlation between stock market and cryptocurrency market |
| Daily Gold Price | Whether any correlation between risk hedge asset and cryptocurrency market |

1. **Computational Setup / Steps**
   1. File name

|  |  |
| --- | --- |
| File name | Purpose |
| Prices.xls | Historical gold Price |
| S&P 500 Index Historical Stock Prices.csv | S&P 500 Index |
| all\_currencies.csv | Cryptocurrency historical report |

* 1. Import relevant libraries

pandas, statsmodels.formula.api, from statsmodels.stats.outliers\_influence import variance\_inflation\_factor, numpy, seaborn, matplotlib.pyplot, dmatrices from patsy, warnings

set=warnings.filterwarnings('ignore')

* 1. Define functions
* Define read\_data() function: the function import and clean the date required for the analysis and return df\_CurrencyMerged dataframe

1. Import csv file ‘all\_currencies.csv’ as data frame df\_currency
2. Import csv file ‘S&P 500 Index Historical Stock Prices.csv’ as data frame df\_SP500
3. Import excel file ‘Prices.xls’, read sheet ‘Daily’, and skip top 8 rows as data frame df\_GoldPrice
4. Merge df\_currency, df\_SP500, and df\_GoldPrice based on Date and return df\_CurrencyMerged data frame with ['Date’, 'Symbol', 'Open\_Price', 'Low', ' SP500ClosePrice ', 'Volume', 'Market\_Cap', Gold\_Price] columns

* Define regressForN(df\_CurrencyMerged,n) function: the function takes df\_CurrencyMerged as input, runs regression based on top n performers, and return regression summary

1. Top n is based on max Market\_Cap of each symbol
2. Regression formula: "Market\_Cap ~ Volume+Open\_Price+SP500ClosePrice+Gold\_Price"

* Define calVIF(dateset, symbol) function: the function takes the dataset and symbol of currency and produce VIF result table

1. Subset the symbol required

2. Run regression using Market\_Cap ~ Volume+Open\_Price+SP500ClosePrice+Gold\_Price

3. Use variance\_inflation\_factor to run VIF calculation and return the result

* Define ccRegress(dataset, symbol) function: the function taken in the dataset and the symbol given and return the correlation coefficient
  1. Market share analysis based on top performers

We found the market share of the cryptocurrencies on 1st January, 2016 and later compared it to market share of cryptocurrencies on 1st January, 2018

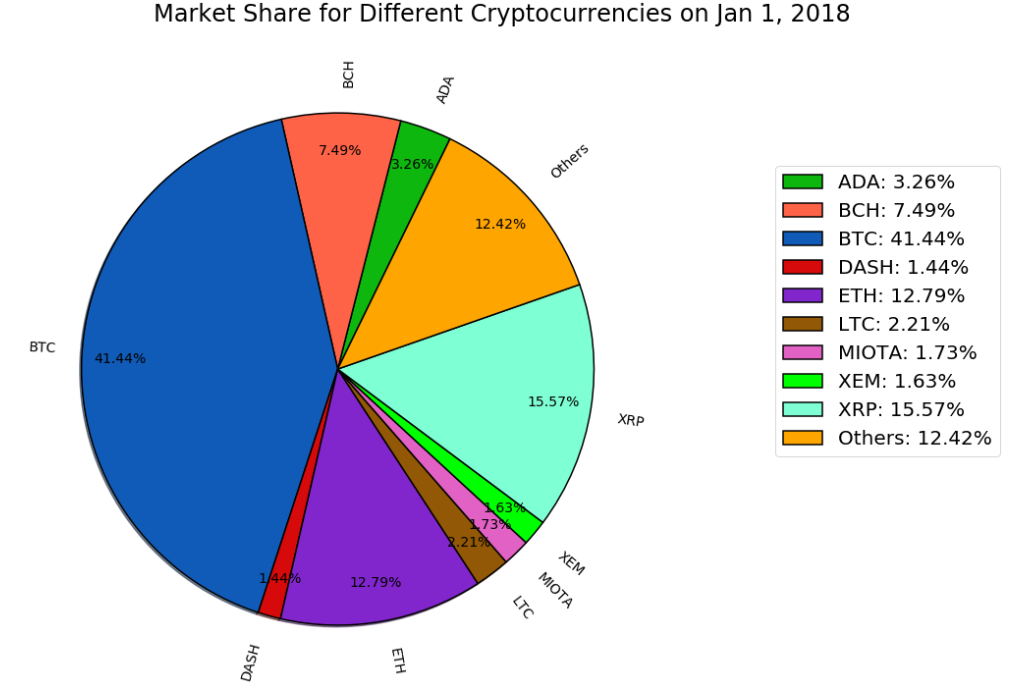
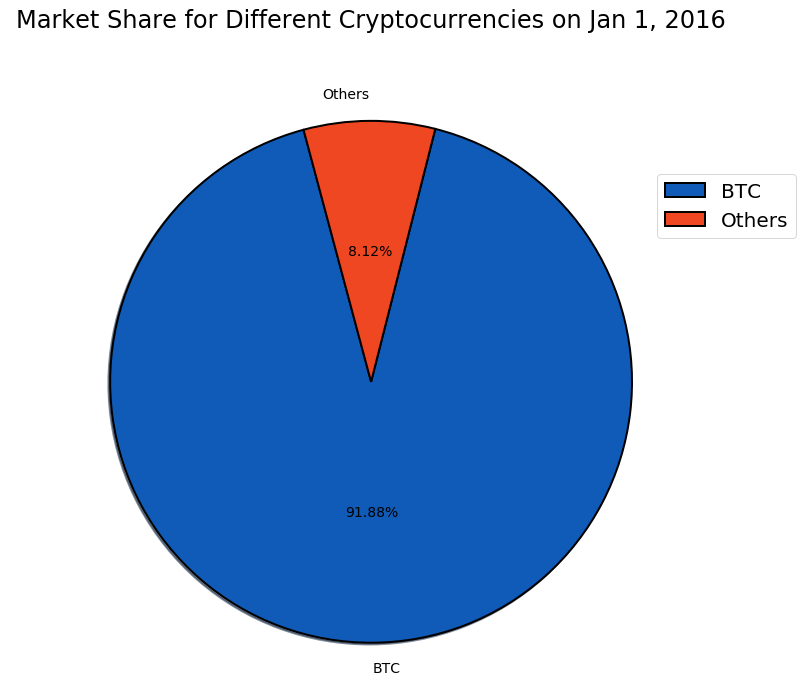
* 1. Normalized Trend Line: We standardized the trend lines to compare on the normal scale the increase in the prices and market cap for cryptocurrency, S&P500 and gold price.
  2. Regression
* Finding regression results for for top n currencies
* Regression summary
* Test VIF

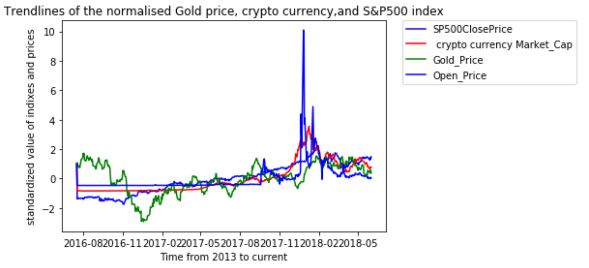
2.7 Clustering: We regress the dependent variables for different currencies and calculate n number of regression equations for n currencies. This would denote the trendlines for these currencies with respect to the given variables i.e. Open price, market cap, gold price and S&P 500. To find the similarity within these trendlines we cluster them using these beta variables of different currencies with K means clustering. We can then prioritize the clusters that has similar trendline as the popular Bitcoin.

* 1. Comparison Analysis
* We analyzed 2 groups of crypto currencies. One group is the mainstream currencies, the other group is the newly introduced currencies.
* Select targets: From the sorted dataframe, we chose 5 types of currencies (except BTC) that have the highest market cap
* Make a line graph to see the trends of these mainstream currencies' prices: Using matplotlib.pyplot library.
* Regression analysis: We did regression analysis to test the relationship between BTC and each one of the mainstream currencies. In this regression model, mainstream currencies’ prices are dependent variables, and BTC’s price is independent variable. Using statsmodels.formula.api library, and using sm.ols function to conduct this analysis. Besides, we also used stats.linregress function and np.corrcoef function, to get the correlation coefficients of each comparisons.
* Then we used the same method to compare the market cap between BTC and other currencies.
* For the second group, we selected 5 new currencies that are recently released and repeated the steps before, make graphs and do regression analysis between the newly introduced currencies and BTC.
  1. Simulation and Prediction
* We used simulation prediction to predict BTC’s price BSM (black-scholes-merton) model.
* We set some parameters, including the current BTC price, option strike price, time cycle, risk neutral payoff rate, rate of fluctuate, number of time steps and simulation times. (During this step, we used np.random.seed function to generate random numbers, and used np.random.standard\_normal function to generate random numbers that follow normal distribution).
* We built a function referring to BSM (black-scholes-merton) model and ran 20000 times simulation to see what are the probabilities of the future price. We used matplotlib.pyplot library to draw the probabilities distribution graph.
  1. Profiling testing, we are using %%time to test the time taken by each block. we observe that the maximum time is taken by the regression of n variables. It is since we are trying to fit curves for 200 variables 200 times and so 200 times the time taken by regression. Reducing the variable used for regression can reduce the time.

1. **Results**
   1. Descriptive analysis

In 2016 BTC owned ~92% market share, however, the picture changed in 2018 when BTC only owns ~42% of the market share followed by XRP (~15%) and ETH (~13%).

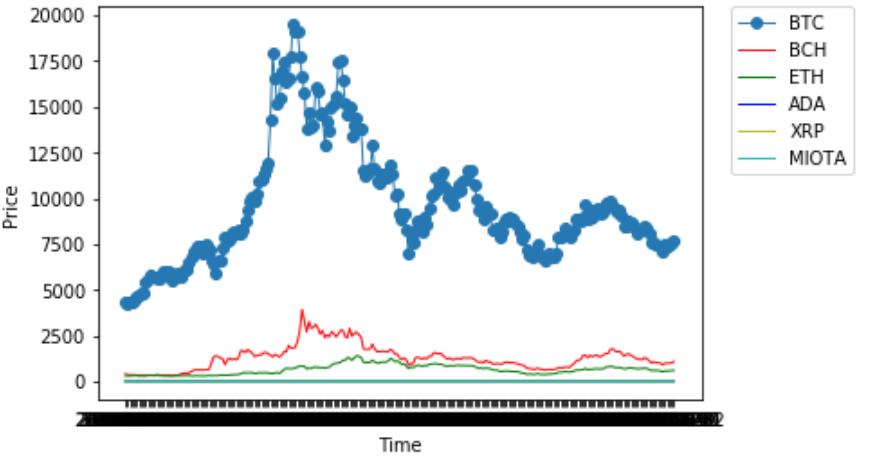
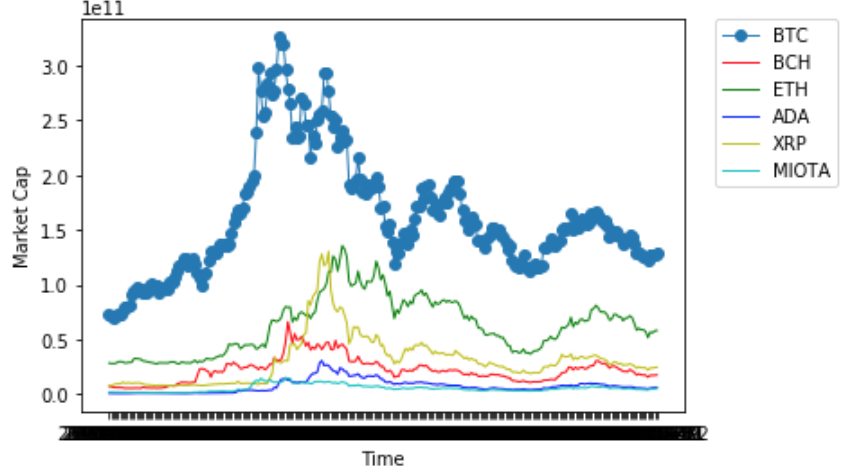
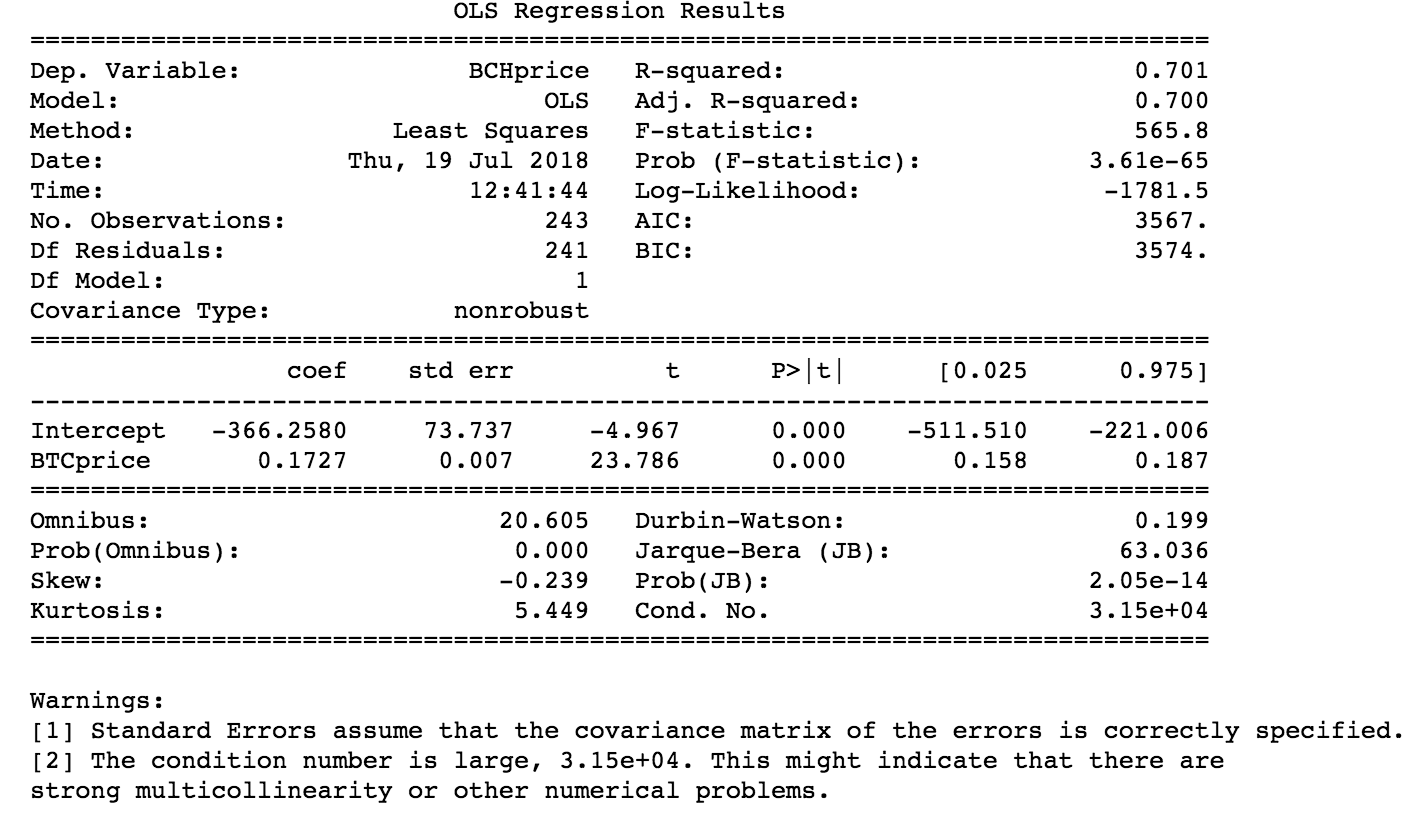


* 1. Normalized trend line: From the trendline we observe that the cryptocurrency curve is the steepest of all, suggesting it has the highest return in the same time frame. We also observe that there are instances where the graph has high peaks, so we have higher returns opportunities.
  2. Regression

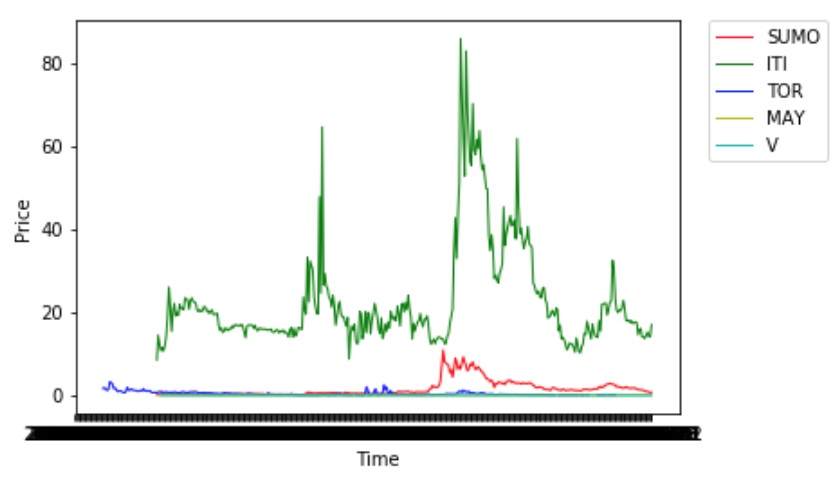
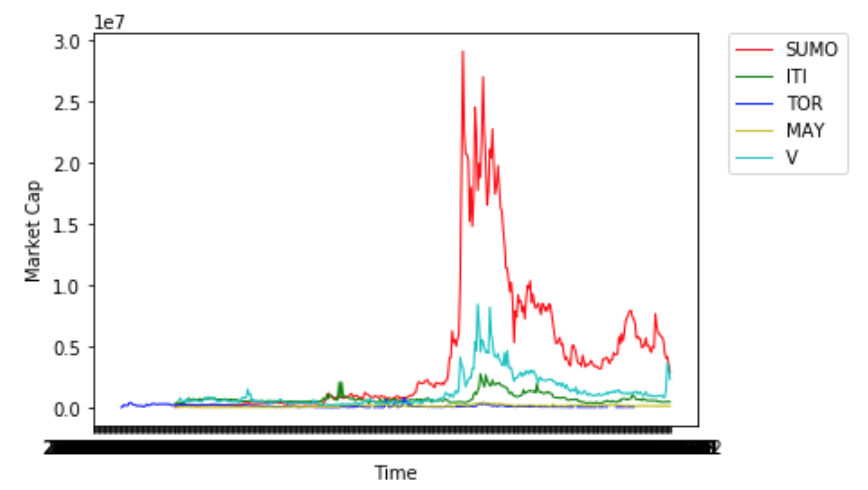
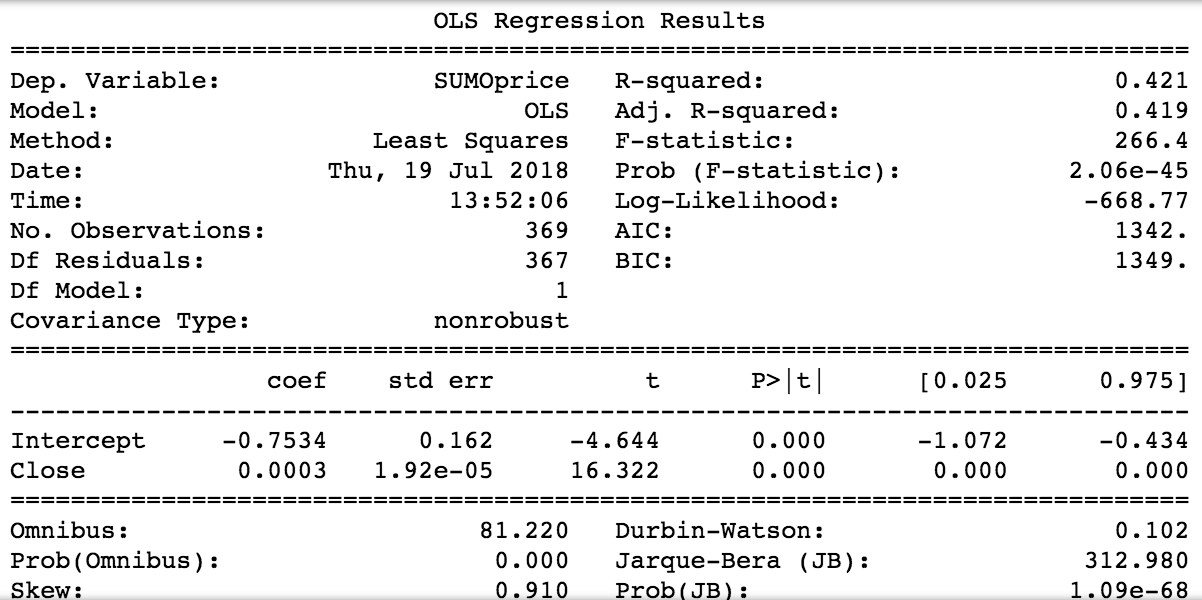
(Ansh)

* 1. Comparison

The first graph below shows the price trends of BTC and other mainstream currencies. The second graph compares the market capital between mainstream currencies and BTC.

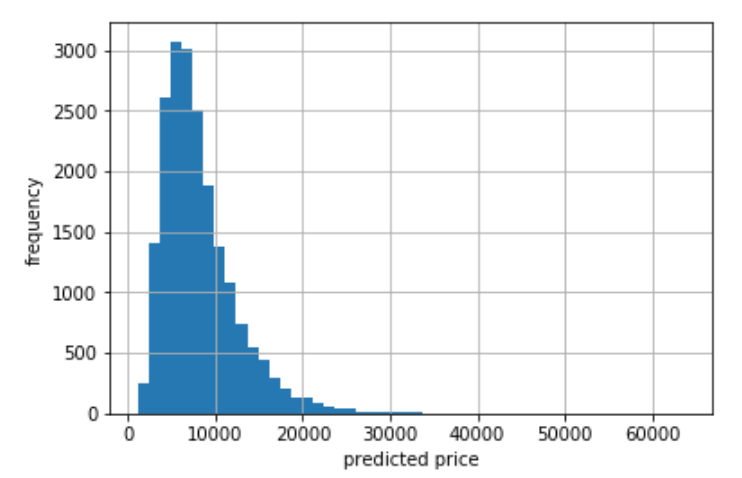
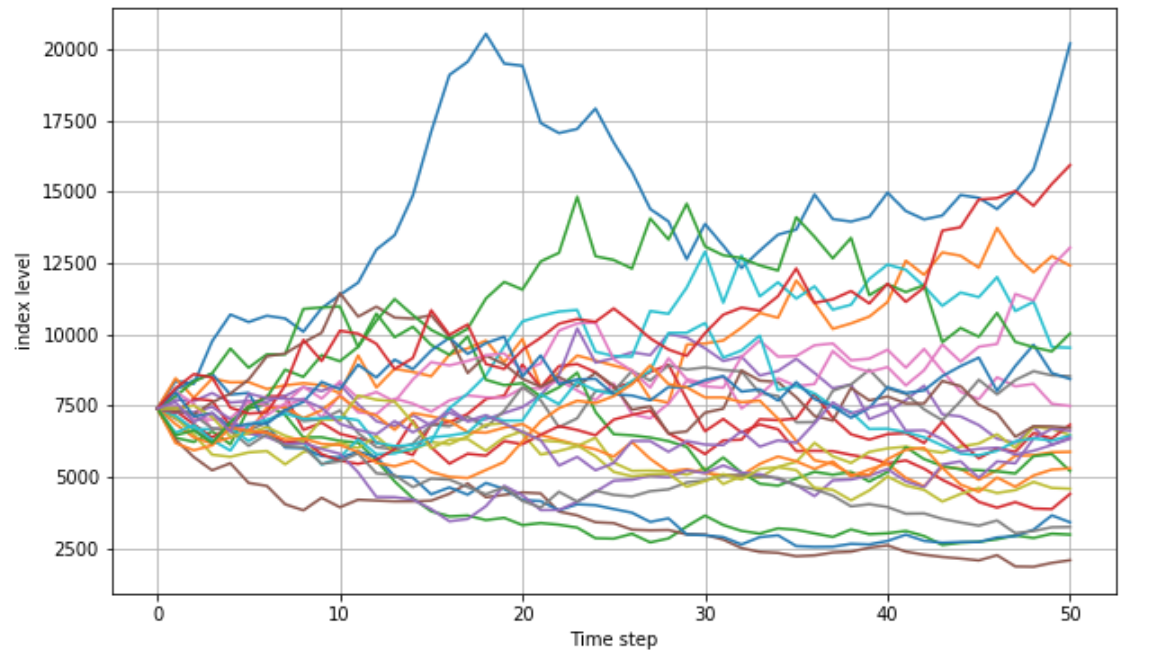
********

The first graph below shows the price trends of BTC and newly-introduced currencies. The second graph compares the market capital between new currencies and BTC.

**** **** 

* 1. Simulation and Prediction

The following graphs are the results of prediction. The first graph shows the probabilities of the top 25 paths from this mode. The second graph shows the frequency of the predicted price. From the graph we can see that the predicted price will fall between 6000 and 8000, and the probability distribution frequency is shown in this histogram.

****

Probabilities of the BTC’s Price

1. **Conclusions**

Daily coin trading volume, daily coin open price, daily S&P 500 index, and daily gold price are all influencing crypto currencies’ market caps. In addition, different currencies highly depend on different variables.

The relationship between mainstream currencies and BTC are very close. Or we can say, the trends of mainstream currencies are following the trend of BTC. However, the relationship between new currencies and BTC are not close, or new currencies’ prices are independent of BTC’s price. Therefore, it is very hard to predict the new currencies' trends based on the trend of BTC’s price or the trend of market.

Profiling conclusion: time taken in regression depends on the number of variable used to regress an output. We have taken top 200 currencies to regress and analyzed the trendline for different currencies. We then do the k means clustering. We also observe that the most time is taken in the plotting the graphs, line pair plot for k means clustering.

1. **Further Study**

* We referred to some research on the modeling and prediction of Bitcoin prices. These materials have mentioned other variables that could be influential to the price of BTC. These factors include S&P 500 index, Dow Jones 30 index, Nasdaq, crude oil, gold, volatility index of the market, etc. Many of these factors haven’t been included in our research. In the further study, we can include more variables that could be influential.
* We used simulation to predict BTC price, however, this model is quite simple, and the result is highly dependent on the parameters we set. In the future research, we can improve our model by simulating the independent variables’ prices (oil price, stock index, which are easier to set parameters, etc), and then use the regression model to predict BTC price.

**Reference**

[1] HUISU JANG AND JAEWOOK LEE 2017. An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information. *IEEE Access(10.1109/ACCESS.2017.2779181), 5433-5434*