Analysis of Bitcoin Prices

I029: Saachi Kotadia **I031: Kashyap Malhotra** **I036: Anubha Mishra** **I049: Shashwat Shankar**

1. Abstract

The paper aims to study the performance of Bitcoin using cryptocurrency elements, commodities and interest rates.

2. Introduction

The system uses multiple machine learning models to analyze historical data of bitcoin prices, along with other financial indicators, such as repo rate, gold prices, bitcoin block size, number of transactions, crude prices etc. and find a relationship between these and the price fluctuations of bitcoin. Based on this analysis, the future price of bitcoin can be predicted.

3. Hypothesis

The initial proposed model is:

btc.price = $\beta 0$ + $\beta 1$(btc.blocksize) + $\beta 2$(btc.txcount) + $\beta 3$(crude.price) + $\beta 4$(gold.price) + $\beta 5$(repo.rate) + ε

4. Methodology

Gathering Data

The data was collated from various sources relating to Bitcoin and other independent variables affecting it: blocksize, number of transactions, gold price, crude oil price, and repo rate.

5. EDA

Data Preparation:

Since some of the data gathered was not available to the lowest level of detail, i.e. daily, the missing values were imputed using the TS package. This data was then split into training and testing sets, that is, data from 2014-2019 for training and 2019-2020 for testing.

Dickey–Fuller test: p=0.436282

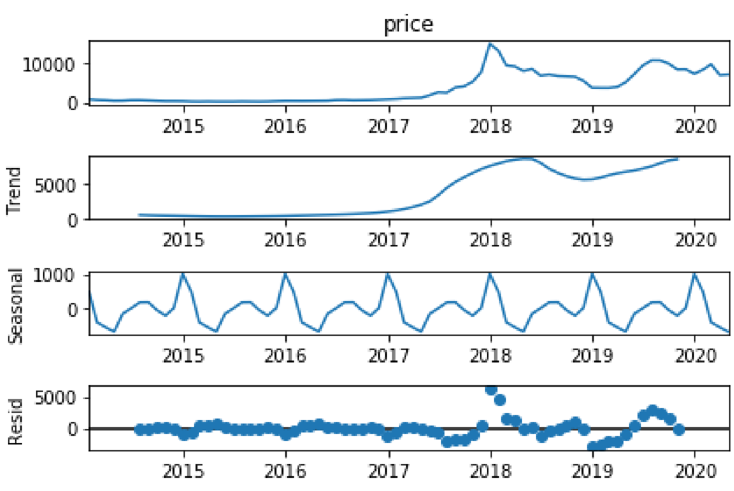

Fig 1 : Stationarity Test

The p value 0.457562 is larger than the critical value, hence we fail to reject the null hypothesis and conclude that the time series is non stationary but has seasonality.

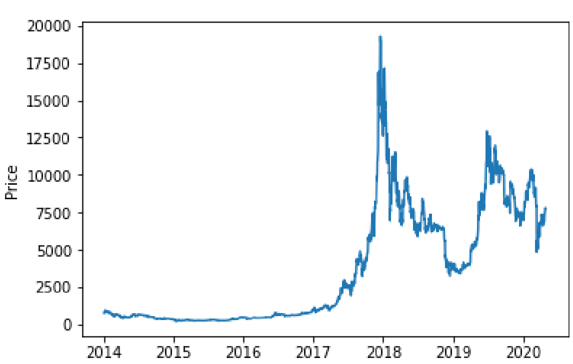

Fig 2 : Bitcoin Prices over the years

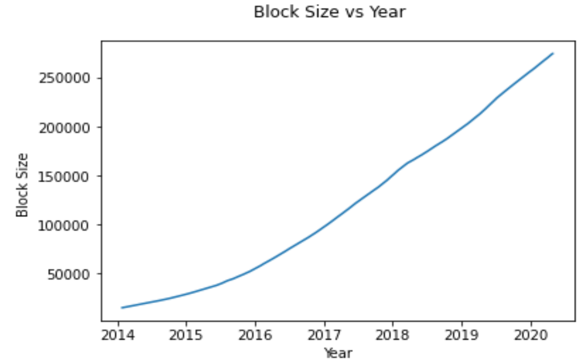

Fig 3 :Increase in Block Size over the years

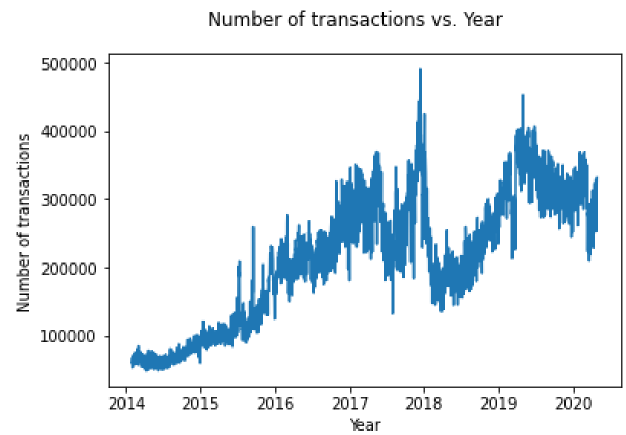

Fig 4 : Number of transactions across the time period

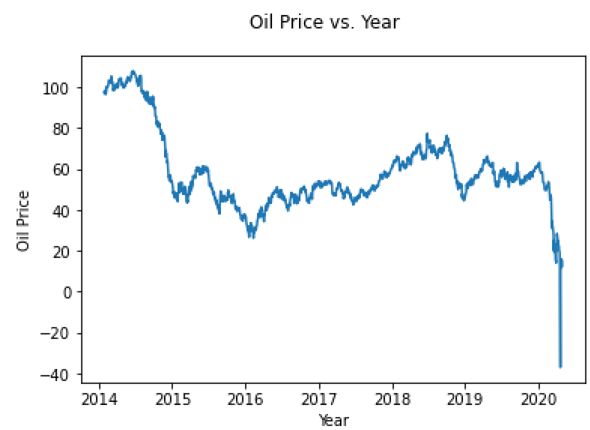

Fig 5 : Oil Prices over the years

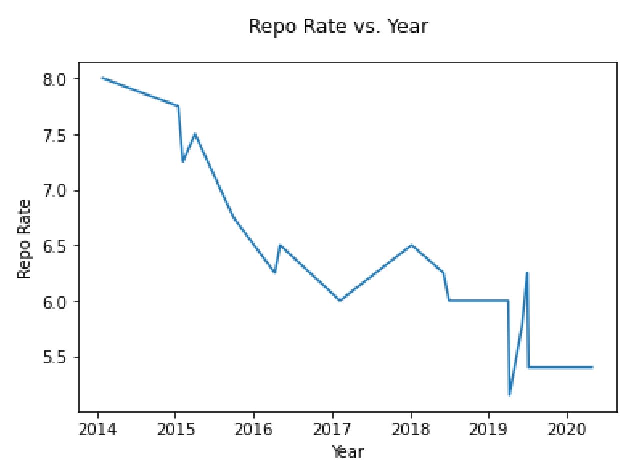

Fig 6 : Repo Rate across the time period

Variable Selection:

Variance Inflation Factor (threshold value = 10) is applied to detect any multicollinearity in the features. The features with VIF>10 were transformed with different transformations.

6. Machine Learning Models

We applied multiple linear regression, ridge regression and lasso regression to our data and compared parameters like r-squared score, AIC, BIC to select the optimum model.

Following machine learning models were used:

Models	**Adjusted R-square**
Multiple Linear Regression	91.31%
Ridge Regression	91.34%
Lasso Regression	91.34%
OLS Linear Model (With transformed variables)	93%

Table 1 : Comparison of Adj R-squared score for different models.

7. Conclusion

OLS Linear Model with transformations enabled us to establish a model whereby all the assumptions were acceptable, VIF scores were less than 10 and all significant variables explained the prices of Bitcoin. Although the model helped us explain the variables and its relationship, it did not work well in predicting the future prices of Bitcoin. The model can be improved in future by including more variables.

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.930
Model:	OLS	**Adj. R-squared:**	0.930
Method:	Least Squares	**F-statistic:**	4798.
Date:	Sat, 24 Apr 2021	**Prob (F-statistic):**	0.00
Time:	16:49:25	**Log-Likelihood:**	-14852.
No. Observations:	1799	**AIC:**	2.972e+04
Df Residuals:	1793	**BIC:**	2.975e+04
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-0.5783	0.008	-74.038	0.000	-0.594	-0.563
txn	-0.0087	0.000	-25.556	0.000	-0.009	-0.008
fee	39.3814	0.479	82.249	0.000	38.442	40.320
search_trends	126.2339	1.532	82.379	0.000	123.229	129.239
index	2.754e-07	1.11e-08	24.861	0.000	2.54e-07	2.97e-07
oil_price	16.0604	1.986	8.086	0.000	12.165	19.956
block_size	2.164e-13	2.75e-14	7.858	0.000	1.62e-13	2.7e-13
repo_rate	-5.8865	0.382	-15.394	0.000	-6.637	-5.137

Omnibus:	589.199	**Durbin-Watson:**	0.200
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	5531.412
Skew:	1.259	**Prob(JB):**	0.00
Kurtosis:	11.213	**Cond. No.**	3.93e+16

Fig 7 : Output of the OLS Linear Model with Transformed Variables

R-squared and Adj. R squared: These values signify the percentage variation in the dependent variable that is explained by the independent variables. Here, 93% variation in the price is explained by the independent variables (gold price, crude oil price, block size etc.). Adj. R-squared is the modified version of R-squared which is adjusted for the number of variables in the regression

Prob(F-statistic): It gives us the overall significance of the regression. As it reaches close to zero, it is implied that the regression is meaningful. In the above figure, it is seen that the Prob(F-statistic) is 0.00, implying that the regression is meaningful.

AIC/BIC: These penalize the errors in case a new variable is added in the equation. Lower AIC implies a better model.

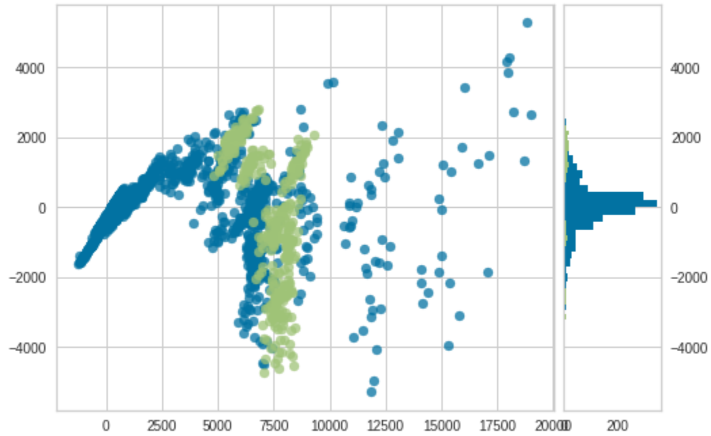

Fig 8 : Residuals v/s Fitted Values for the Linear Regression Model

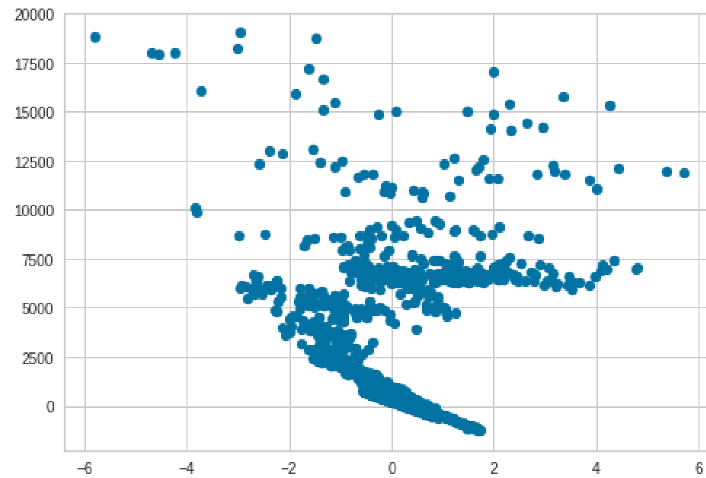

Fig 9 : Residuals v/s Fitted Values for the OLS Linear Model with Transformed Variables

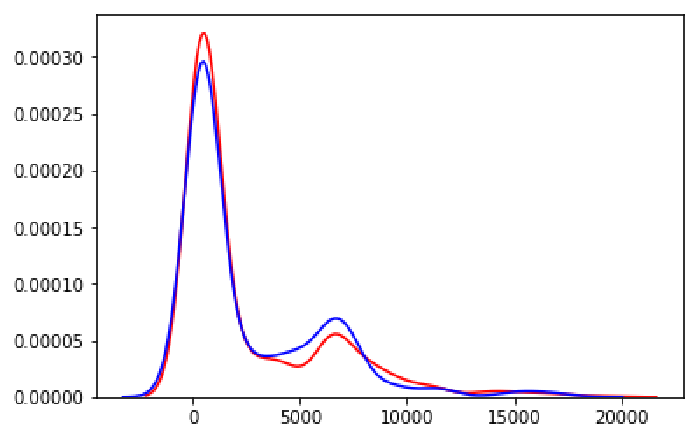

Fig 10 : Predicted v/s Actual Bitcoin Prices

8. Future Work

Important data like economic policies, social media events could be included for a more accurate fit.