Abdullah Mobeen

Professor Jeffrey Simonoff

Regression and Multivariate Data Analysis

NYU | STERN

"I think the internet is going to be one of the major forces for reducing the role of government. The one thing that's missing but that will soon be developed, is a reliable e-cash." This quote by Professor Milton Friedman, a Nobel Prize winner in economics, describes what a cryptocurrency is. Recently, we have observed a massive increase in the value of the most famous cryptocurrency – Bitcoin. What followed is this aura of a sudden interest in bitcoins. While it is true that the general motivation to get involved in this cryptocurrency – bitcoin, is to make monetary profit, the driving force behind Bitcoin is still its fascinating technology. In this report, I'm going to discuss the regression model for the data related to Bitcoin and see if I find anything interesting that helps predict the value of Bitcoin.

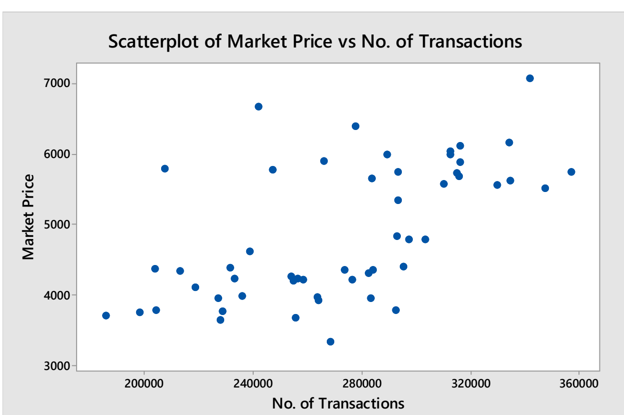
Before laying out the dynamics of my regression, let me briefly explain what a Bitcoin is. Bitcoin is one of the many cryptocurrencies that are available to the public for trading and to understand Bitcoins, it is vital to understand the technology behind Bitcoins or generally behind cryptocurrencies. The foundational idea is of blockchain technology that is based on a peer-to-peer infrastructure. This means that there is no central authority when it comes to blockchain technology and that means there is no single point of failure. Even if one of the 'peers' is experiencing a failure, others stay stable. This idea is really popular in the domain of Computer Networking where we often cannot afford to have a single point of failure, as every user connected to a particular server would be disconnected due to a single failure in the network infrastructure. So now if we think about the sectors that manage money in a given country – Government and Banks, we realize that these sectors are highly centralized. This centralized nature of monetary institutions attracts people with intentions to hack systems and steal assets because attacking one particular system would allow them to affect all the users connected to it. This is what blockchain technology and more specifically, cryptocurrencies aim to tackle. So what is actually fascinating about Bitcoin, other than its increasing monetary value, is its ability to ensure fraud-free transactions and maintain an authentic level of security based on peer-to-peer infrastructure, not on centralized infrastructure.

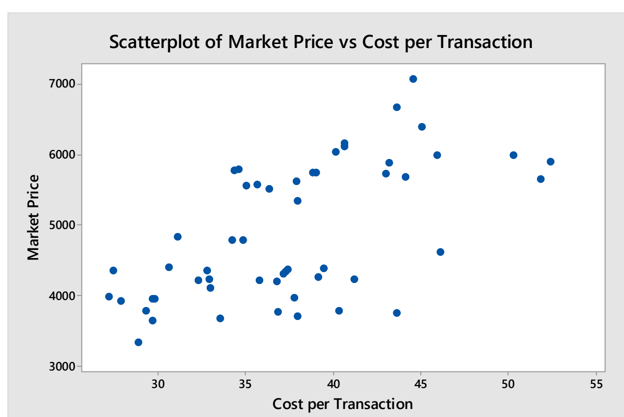
Having explained what a Bitcoin is and why we should look forward to it, let me begin with my analysis. I have tried to build a model that can give us some insight on how to realize the trend in the value of a bitcoin. So my **response variable** is the **Market Price of Bitcoin (USD).** In order to understand the trend in the market price of Bitcoin, we need to realize that Bitcoin is traded as a commodity and hence its demand and supply play a vital role in estimating its market price. It is also noted that the more Bitcoins are reported in news, the more it's demand increases. So I have collected data for predictor variables that fall under these two categories - Demand and Supply. Predictors estimating the demand for Bitcoins are **no. of Transactions per Day and no. of Google Searches mentioning Bitcoin.** Predictors estimating the supply of Bitcoins are **Total Miners' Revenue (USD), Difficulty,** and **Cost per Transaction (USD).**

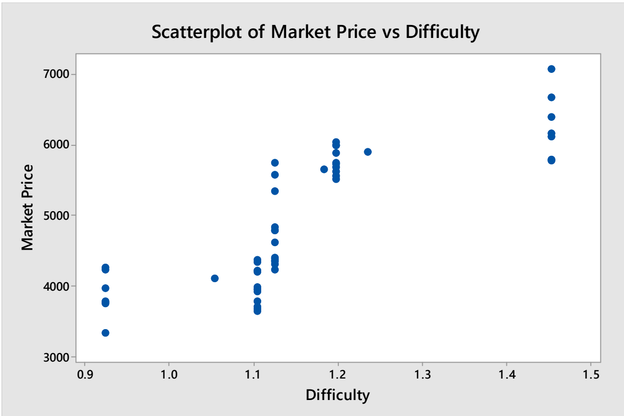
I have collected Bitcoin data for these variables from 10th September 2017 to 1st November 2017, in total 53 data points. The reason I chose this period is that this period witnessed many fluctuations in the market price of Bitcoins. I found the relevant data through www.blockchain.info, it provides complete day-to-day information on Bitcoins with interactive charts that help visualize the general trends. For the No. of Google Searches mentioning Bitcoin, I acquired the data from Google Trends, as it allows us to search a particular word and see a number of searches over a specified time period. So I collected the No. of Google Searches corresponding to the same time period i.e. 10th September to 1st November.

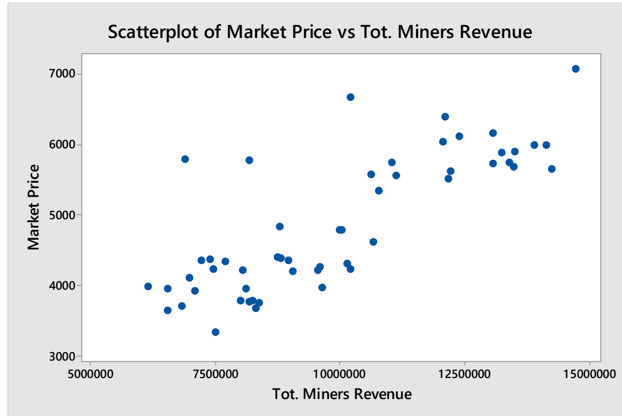
No. of Transactions per Day indicates the number of daily confirmed Bitcoin transactions. **No. of Google Searches** indicates the general online popularity of Bitcoins by giving a percentage of searches on a specified day against the maximum number of searches (with Bitcoin) per day. For example, if the most number of searches for Bitcoin per day is 10, another day with 2 bitcoin searches would be (2/10)*100 percent. To understand the meaning of each predictor that falls under supply of Bitcoin, let us briefly go through the process through which Bitcoins are created. People can create Bitcoins from scratch and that process is known as Bitcoin Mining. It is a process of solving complex algorithms that require a tremendous amount of memory and time on computers to be solved. It is a costly procedure to implement, but once somebody successfully solves those algorithms, they 'mine' Bitcoin in shape of reward and their profit is the cost of mining subtracted from the value of Bitcoin rewarded. So **Total Miners' Revenue** indicates the total value of transaction fees paid to miners. **Difficulty**, as indicated by the name, is the measure of how difficult it is to mine bitcoins. And **Cost per Transaction** is Total Miner's Revenue divided by Number of Transactions. The last three predictors are often considered by people while deciding whether to mine Bitcoins or not, hence they give a good estimate of the supply of Bitcoins.

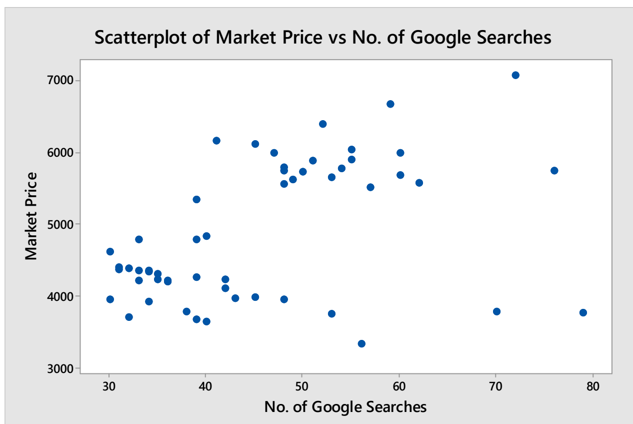
Let us look at the individual scatterplots of the response variable versus each predictor to get a rough understanding of what the individual relations look like.



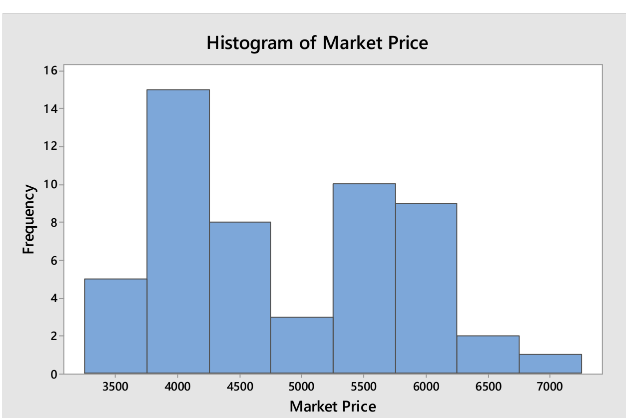


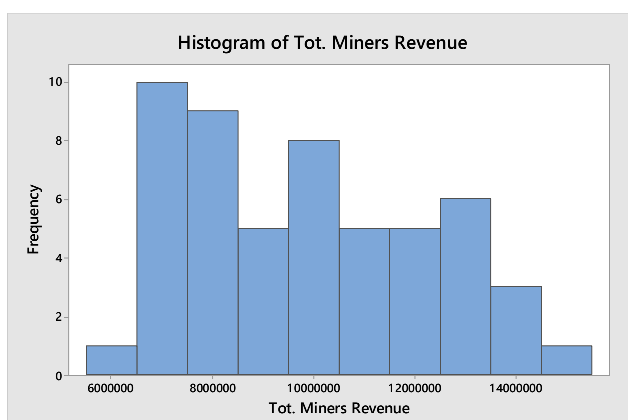


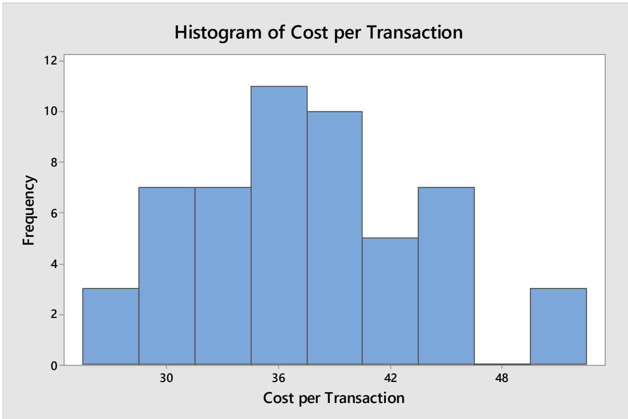


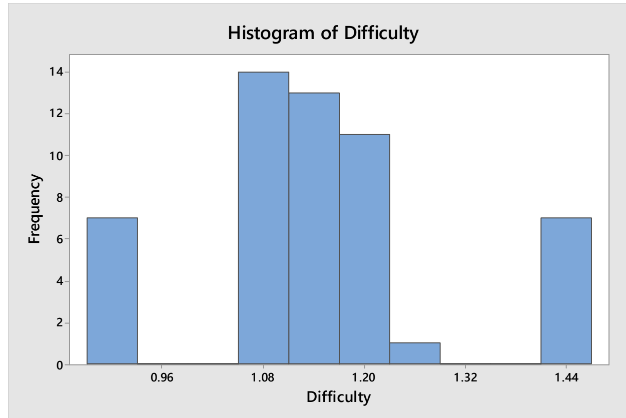


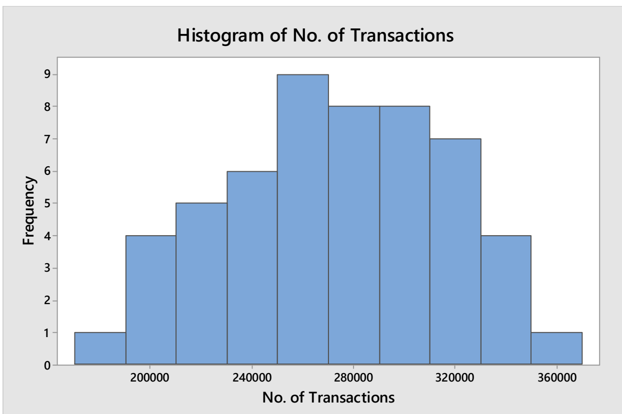
We can notice some sort of correlation existing between our response variable and each of our predictors. Let us now take a look at the histograms of the response and predictor variables:

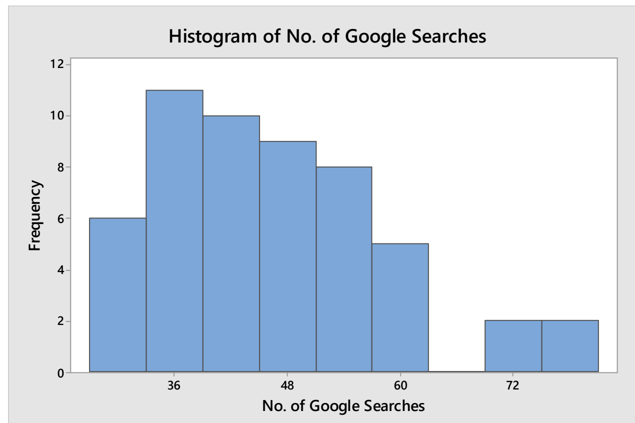




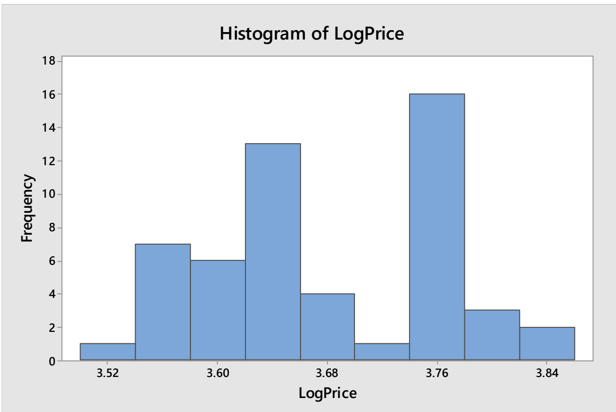

Histogram of Cost per Transaction
Frequency
Cost per Transaction
0
2
4
6
8
10
12
30
36
42
48

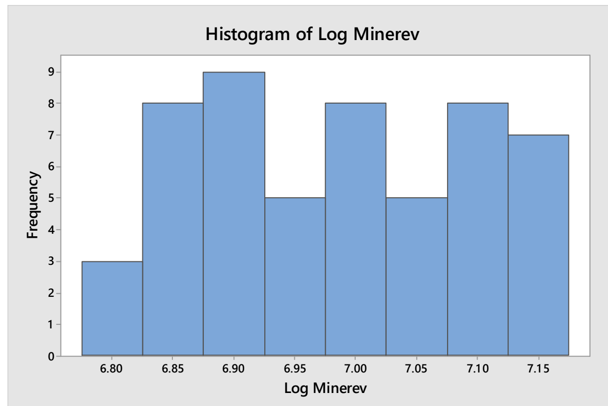

Histogram of Difficulty
Frequency
Difficulty
0
2
4
6
8
10
12
14
0.96
1.08
1.20
1.32
1.44

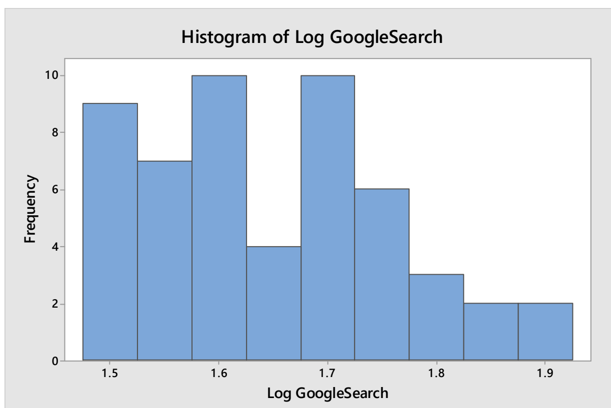

Histogram of No. of Transactions
Frequency
No. of Transactions
0
1
2
3
4
5
6
7
8
9
200000
240000
280000
320000
360000


Histogram of No. of Google Searches
Frequency
No. of Google Searches
0
2
4
6
8
10
12
36
48
60
72

It is not surprising for us to see the variables describing money having long right tailed histograms. So to correct for them we will take the log of Market Price, Total Miners' Revenue, and No. of Google Searches (non-monetary variable). The histograms of these variables after taking logs are as follows:







We have, to some extent, accounted for the long right tailed histograms. The result of the preliminary regression is as follows:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	0.336436	0.067287	85.82	0.000
No. of Transactions	1	0.000068	0.000068	0.09	0.770
Cost per Transaction	1	0.000085	0.000085	0.11	0.744
Difficulty	1	0.097387	0.097387	124.22	0.000
Log Minerev	1	0.000866	0.000866	1.10	0.299
Log GoogleSearch	1	0.004565	0.004565	5.82	0.020
Error	47	0.036849	0.000784		
Total	52	0.373284			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0280002	90.13%	89.08%	87.60%

Coefficients

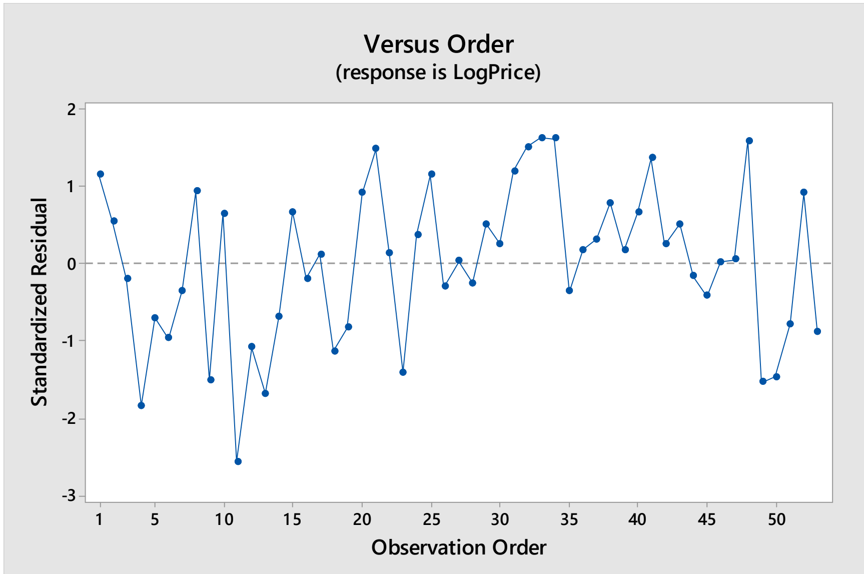
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0.70	3.33	-0.21	0.834	
No. of Transactions	-0.000000	0.000001	-0.29	0.770	96.56
Cost per Transaction	-0.00207	0.00629	-0.33	0.744	99.10
Difficulty	0.3340	0.0300	11.15	0.000	1.24
Log Minerev	0.571	0.544	1.05	0.299	221.70
Log GoogleSearch	0.0956	0.0396	2.41	0.020	1.29

Regression Equation

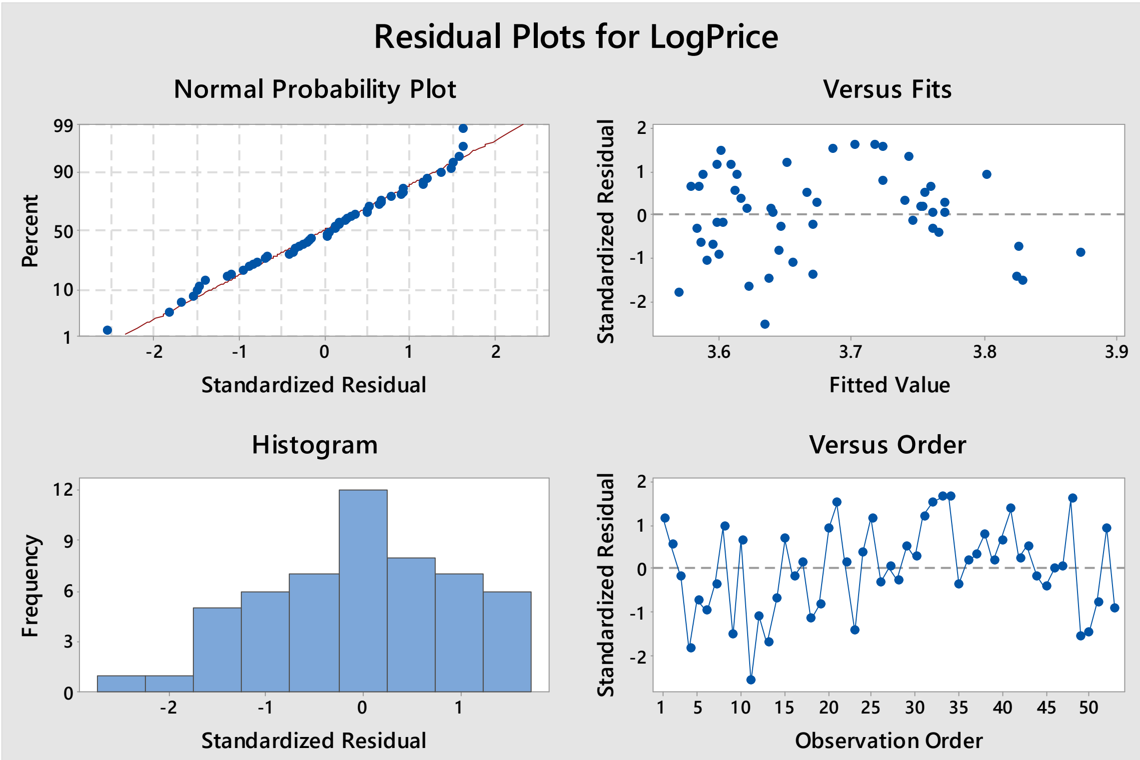
LogPrice = -0.70 - 0.000000 No. of Transactions - 0.00207 Cost per Transaction
+ 0.3340 Difficulty + 0.571 Log Minerev + 0.0956 Log GoogleSearch

Although we notice a high R-sq of 90.13% and a few redundant predictor variables, this model suffers from a difficulty. Since it is a time series data, we need to address the time ordering of data and check for autocorrelation. When considering the methods to identify autocorrelation, four main methods stand popular – Checking the residuals versus order graph, The Durbin-Watson statistic (highly parametric), ACF Plot (Semi-Parametric), and Runs Test (Non-

Parametric). We will conduct all these tests to see if we can identify any autocorrelation in our data. To begin with, let's take a look at the residuals versus time-order graph to see if we're able to detect any autocorrelation from it.



The Standardized Residuals versus Order plot does not really help in identifying autocorrelation. We could see a weak pattern around residuals from 30 to 45. This is not sufficient in declaring that we have autocorrelation so we will also conduct the other three tests in order to check what they indicate. Let's use the DW statistic but in order to consider the DW statistic valid, we need to make sure that we have satisfied all the fundamental assumptions of Regression i.e. all the errors are independent, identically distributed, and normal with mean 0. Let's have a look at other residual plots to validate our assumptions.

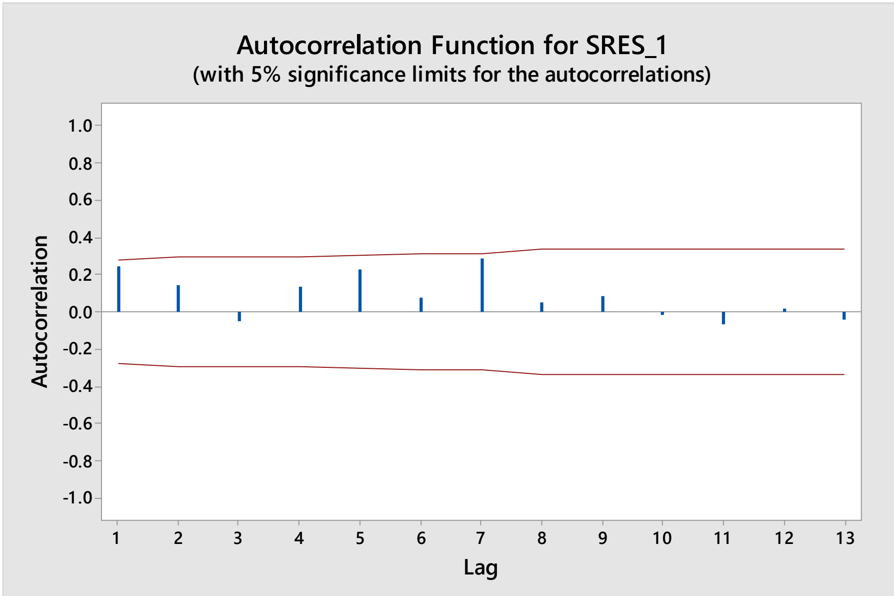


We could see that the errors are normally distributed with a few outliers towards the top right. In our standardized residuals versus fitted plot, we can identify a cyclic relationship that does not validate our assumption of homoscedasticity. The Durbin-Watson statistic is as follows:

Durbin-Watson Statistic

Durbin-Watson Statistic = 1.45374

The Durbin Watson statistic for our model falls between the critical values of D_L (1.209) and D_U (1.592) for n = 53 and p = 5. However, since our model does not completely stay true to all the regression assumptions, we cannot be really confident in what our Durbin Watson statistic indicates. So let's try the semi-parametric test – ACF plot, to see whether it indicates autocorrelation in our model or not. The ACF plot is used to estimate the true underlying autocorrelations among the data points and does not make many assumptions about the model like the Durbin-Watson statistic.



Autocorrelations

Lag	ACF	T	LBQ
1	0.244494	1.78	3.35
2	0.140783	0.97	4.48
3	-0.047476	-0.32	4.62
4	0.135153	0.91	5.70
5	0.232155	1.54	8.97
6	0.078889	0.50	9.36
7	0.291085	1.84	14.73
8	0.049599	0.30	14.89
9	0.083115	0.49	15.35
10	-0.020413	-0.12	15.38
11	-0.070441	-0.42	15.72
12	0.021014	0.12	15.75
13	-0.043020	-0.25	15.89

The ACF plot does not show any autocorrelation as all the bars on all 13 lags are statistically insignificant at 0.5 level. The first order autocorrelation is 0.244, which is not very large. Our ACF plot also does not support the AR(1) process as there is no geometric decay in the bars with increasing number of lags. The autocorrelation also does not exhibit the exponential decay of 0.244^1 ,0.244^2 , 0.244^3, … , 0.244^{13}. We kept the maximum number of lags to 13 ~ n/4, as our dataset is small and increasing the number of lags would keep deleting the datapoints from our dataset.

Finally, we would also conduct a non-parametric test – Runs Test to see if that indicates anything about the autocorrelation in our dataset.

Runs Test: SRES_1

Descriptive Statistics

		Number of Observations	
N	K	≤ K	> K
53	0	23	30

Test

Null hypothesis	H_0: The order of the data is random
Alternative hypothesis	H_1: The order of the data is not random

Number of Runs		
Observed	Expected	P-Value
24	27.04	0.391

Although the Runs Test shows a slight positive autocorrelation, it is not statistically insignificant with a p-value of 0.391, which does not provide enough evidence for us to reject the null hypothesis. So from the four methods above, we could see that our model does not show any strong signs of autocorrelation. This means that now we can continue with our regression and perform the required steps to fit our model better.

We began with five potential predictor variables and in our preliminary multiple regression, we noticed how some of the predictor variables were not adding anything to the predictive power of our model. Let's perform best subset regression on our data to filter out the predictor variables that make our model simple yet powerful in terms of predictive power.

Response is LogPrice

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	No. of Transactions	Cost per Transaction	Difficulty	Log GoogleSearch	Log Minerev
1	62.7	62.0	60.3	128.7	0.052261			X		
1	62.2	61.5	59.4	130.8	0.052568					X
2	88.9	88.4	87.4	5.9	0.028808			X		X
2	77.3	76.3	74.9	61.3	0.041210	X		X		
3	90.1	89.5	88.4	2.1	0.027466			X	X	X
3	88.9	88.2	87.1	7.8	0.029074	X		X		X
4	90.1	89.3	88.0	4.1	0.027732		X	X	X	X
4	90.1	89.3	88.0	4.1	0.027739	X		X	X	X
5	90.1	89.1	87.6	6.0	0.028000	X	X	X	X	X

Here we could see that when we use a three predictor model with Difficulty, Log GoogleSearch, and Log Minerev, we obtain a model with the Mallows Cp less that (p + 1) i.e. Mallows Cp of 2.1 < (3 + 1). The R^2 , R^2_{adj} and R^2_{pred} are all at their maximum with the values of 90.1%, 89.5%, and 88.4% respectively. So let's do a multiple regression only using these three predictor variables.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	0.336319	0.112106	148.61	0.000
Difficulty	1	0.099630	0.099630	132.07	0.000
Log Minerev	1	0.065364	0.065364	86.65	0.000
Log GoogleSearch	1	0.004530	0.004530	6.00	0.018
Error	49	0.036965	0.000754		
Total	52	0.373284			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0274660	90.10%	89.49%	88.35%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.336	0.269	1.25	0.218	
Difficulty	0.3323	0.0289	11.49	0.000	1.20
Log Minerev	0.4018	0.0432	9.31	0.000	1.45
Log GoogleSearch	0.0937	0.0383	2.45	0.018	1.25

Regression Equation

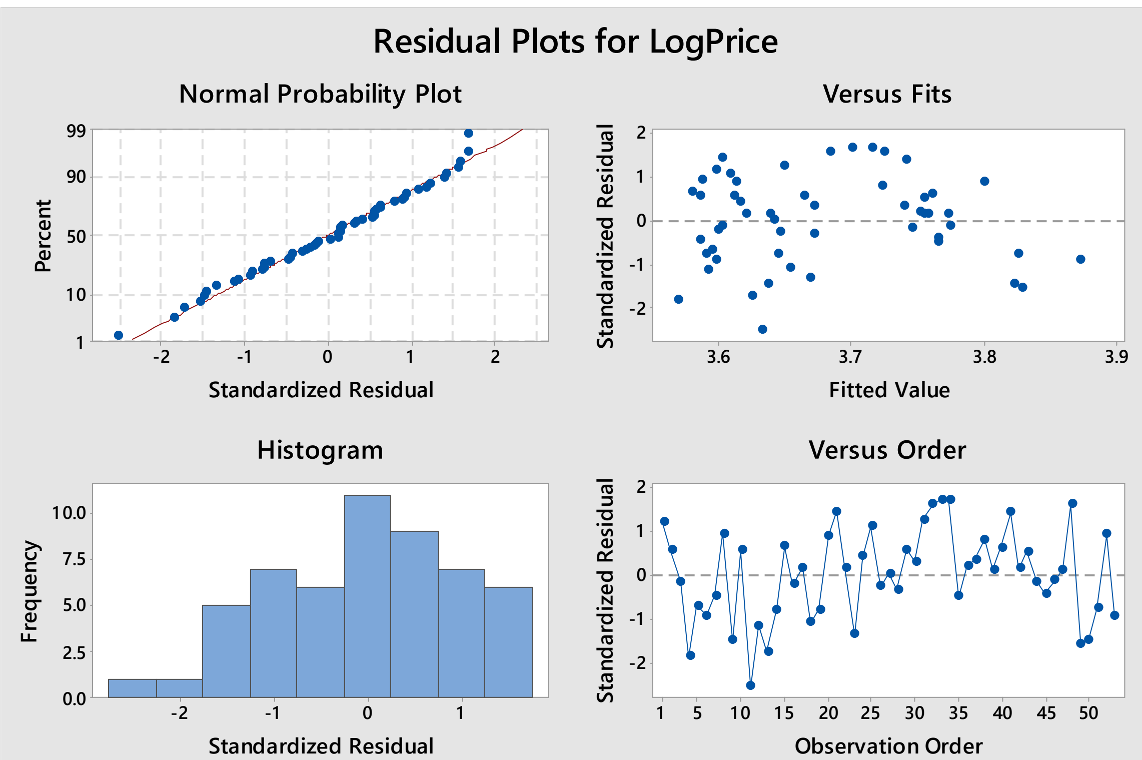
LogPrice = 0.336 + 0.3323 Difficulty + 0.4018 Log Minerev + 0.0937 Log GoogleSearch

Durbin-Watson Statistic

Durbin-Watson Statistic = 1.42864

We get a model with a high R^2 of 90.10%. It shows an error of 0.274660, however we should consider using a corrected measure of standard error i.e. 0.274660 * sqrt((53-3-1)/(53-5-1)) = 0.274660 * 1.02 = 0.28. So we could make a 95% rough prediction in an interval $\pm$2*(0.28) of the Logged Market Price (27% of the Market Price to 3.63 times the Market Price), which translates to 4% wider prediction interval. The regression is statistically significant with a F value of 148.6. Looking at the predictor variables, we can also see that all of them are statistically significant with low p-values corresponding to significant t-values. The regression equation indicates that 1-unit increase in Difficulty corresponds to $100*(10^{0.3323} - 1) = 1.14$ time the market price increase in the market price, holding all else fixed in the model. For the Miners' Revenue, it indicates a 1% increase in Miners' Revenue corresponding to 0.4018% increase in the market price, holding everything else fixed in the model. And a 1% increase in the No. of Google Searches corresponds to 0.0937% increase in the market price of Bitcoin, again holding everything fixed in the model.

Now let's take a look at the residual plots to see if they validate our fundamental assumptions of normality, constant variance, independence etc.

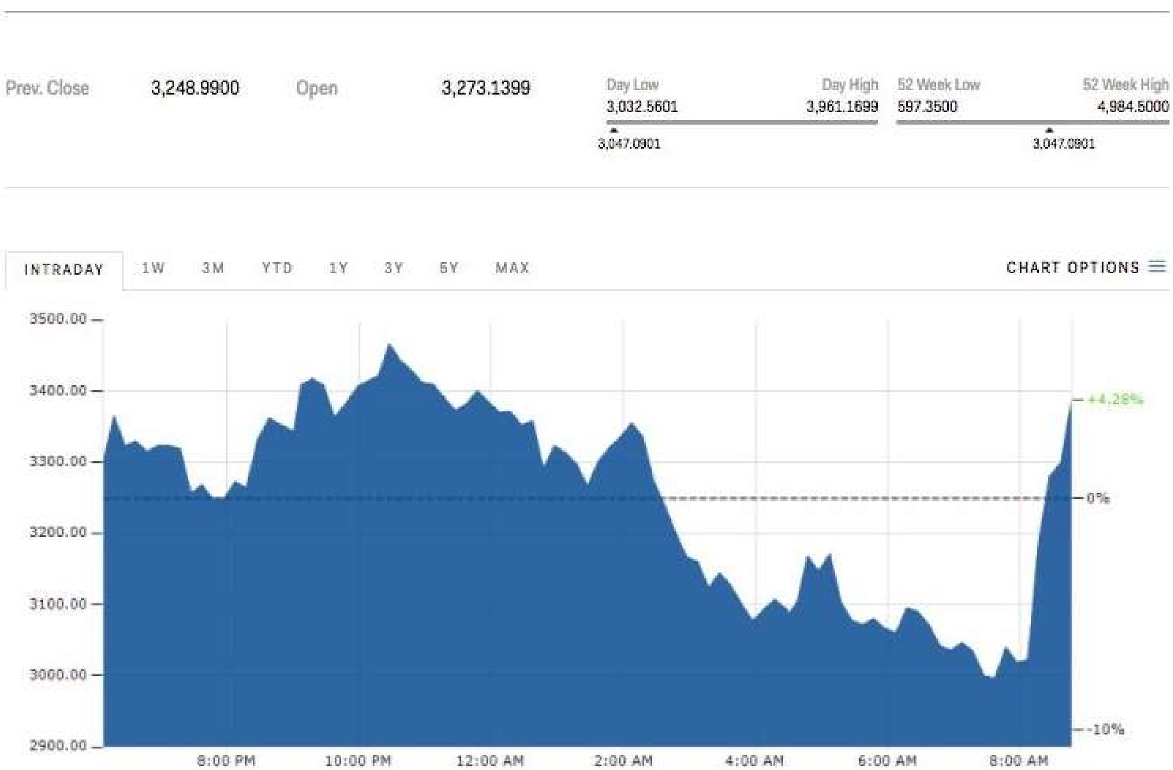


In the normal probability plot, we see straight line that indicates the normality in our data points. There are, however, a few outliers in the plot that we will investigate once we conduct the diagnostics on our dataset. As for the residuals versus fits plot, there is a cyclic pattern that does not support our assumption of constant variance in errors. In order to make our model represent the situation truly, it is essential to address the outliers so let's take a look at the diagnostics:

Date	Row	SRES_3	HI	COOK
9/10/2017	1	1.178338	0.082414	0.031177
9/11/2017	2	0.549115	0.087639	0.007241
9/12/2017	3	-0.14482	0.0772	0.000439
9/13/2017	4	-1.84263	0.115878	0.111251
9/14/2017	5	-0.69527	0.170353	0.024815
9/15/2017	6	-0.91491	0.224541	0.060595
9/16/2017	7	-0.46873	0.089278	0.005385
9/17/2017	8	0.930112	0.056457	0.012941
9/18/2017	9	-1.48925	0.038115	0.021971
9/19/2017	10	0.547245	0.105817	0.00886
9/20/2017	11	-2.52316	0.027229	0.044551
9/21/2017	12	-1.13683	0.071857	0.025014
9/22/2017	13	-1.74037	0.031435	0.024576
9/23/2017	14	-0.79061	0.067565	0.011323
9/24/2017	15	0.629118	0.084636	0.009149
9/25/2017	16	-0.22779	0.055009	0.000755
9/26/2017	17	0.150873	0.045093	0.000269
9/27/2017	18	-1.08768	0.03617	0.011099
9/28/2017	19	-0.78385	0.032827	0.005213
9/29/2017	20	0.891559	0.043947	0.009135
9/30/2017	21	1.415947	0.061113	0.032625
10/1/2017	22	0.150592	0.05704	0.000343
10/2/2017	23	-1.3573	0.046161	0.022289
10/3/2017	24	0.397268	0.045889	0.001898
10/4/2017	25	1.081044	0.053274	0.016441
10/5/2017	26	-0.26644	0.04525	0.000841
10/6/2017	27	0.015994	0.050465	3.4E-06
10/7/2017	28	-0.3265	0.094993	0.002797
10/8/2017	29	0.564164	0.056219	0.00474
10/9/2017	30	0.302727	0.028646	0.000676
10/10/2017	31	1.234431	0.022777	0.008879
10/11/2017	32	1.572632	0.037938	0.024382
10/12/2017	33	1.681884	0.055834	0.04182
10/13/2017	34	1.678346	0.111925	0.088753
10/14/2017	35	-0.48965	0.072293	0.004671
10/15/2017	36	0.173768	0.049902	0.000396

10/16/2017	37	0.330506	0.037372	0.00106
10/17/2017	38	0.80132	0.025781	0.004248
10/18/2017	39	0.125904	0.057031	0.00024
10/19/2017	40	0.618979	0.068739	0.00707
10/20/2017	41	1.402884	0.039783	0.020385
10/21/2017	42	0.14488	0.074102	0.00042
10/22/2017	43	0.519737	0.052614	0.00375
10/23/2017	44	-0.16142	0.044253	0.000302
10/24/2017	45	-0.43367	0.064604	0.003247
10/25/2017	46	-0.13491	0.056895	0.000274
10/26/2017	47	0.119481	0.184453	0.000807
10/27/2017	48	1.570026	0.236678	0.191075
10/28/2017	49	-1.55177	0.124205	0.085375
10/29/2017	50	-1.45765	0.107457	0.063951
10/30/2017	51	-0.77162	0.104353	0.017343
10/31/2017	52	0.902103	0.136785	0.032238
11/1/2017	53	-0.93513	0.151712	0.039098

In the diagnostics above, we could see two potential outliers with high Cook's Distance. These two correspond to row 4 with a value of 0.11 and row 48 with a value of 0.19. Row 4 corresponds to 13th September 2017 and this was the day when Bitcoin suffered a drastic depreciation in market price, falling from a consistent run of $4,300+ to $3,320. Major news outlets also released several reports on 11th and 12th September, speculating how China would impose a ban on Bitcoins, which it eventually did on 27th September. This caused many customers in China to take a step back and not invest in Bitcoins generously. Since the global market is still skeptic of Bitcoins, any news adding to the skepticism has a noticeable impact on the value of Bitcoin. Investigating the second case, we find out that it corresponds to 27th October, 2017. What's unusual about this date is the sudden fall in the Miners' Revenue from $13,483,500 on 25th October to $6,888,711 on 27th October. This provided less motivation to people mining Bitcoins. And since mining Bitcoins is an expensive job, less monetary profits demotivates the miners to continue mining. To further investigate the reason due to which we observe this sudden fall in miners' revenue, we find out that on 26th October MarketWatch reported Warren Buffet calling Bitcoin "a real bubble". Such speculations coming from one of the top investors in the market added to the skepticism towards Bitcoins and we saw that people once stopped investing generously in Bitcoins. As for the leverage points, all the values fall under the limit of 2.5*(p+1/n) = 0.19, except for one – on row 6. There is a leverage point there and investigating that point further reveals to us a sudden rise in the value of Bitcoin. The image below shows this rise on Friday 15th September and is obtained from The Business Insider.



Now that we have identified the outliers and leverage points, what can we do about them? Since we're using a time-series data, we cannot just remove the data points. What we could do as an easy solution is to check the unusual variable values, remove them, and impute new values. These new values are going to be the mean of preceding and succeeding values of that specific row.

After doing so, we essentially have a new dataset. So it is not reasonable to keep using our three predictor model. Let's do the best subset regression again to see what the best model turns out to be.

Response is LogPrice

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	No. of Transactions	Cost per Transaction	Difficulty	Log MinerRev	Log GoogleSearch
1	70.5	69.9	68.6	103.3	0.045409				X	
1	61.6	60.9	59.3	149.0	0.051769			X		
2	89.1	88.6	87.9	9.3	0.027891			X	X	
2	77.4	76.5	75.1	69.5	0.040110	X		X		
3	90.6	90.0	89.0	3.7	0.026191			X	X	X
3	89.4	88.7	87.9	9.7	0.027766	X		X	X	
4	90.8	90.1	88.9	4.4	0.026106	X		X	X	X
4	90.6	89.9	88.7	5.3	0.026351		X	X	X	X
5	90.9	89.9	88.4	6.0	0.026273	X	X	X	X	X

Again it turned out that a simple yet reasonable model is the one with the same three predictor variables with a R^2 of 90.6% and an error of 0.026. Let's look at the multiple regression with these three predictors to see what results it yield.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	0.322639	0.107546	156.78	0.000
Difficulty	1	0.068397	0.068397	99.71	0.000
Log Minerev	1	0.061482	0.061482	89.63	0.000
Log GoogleSearch	1	0.005281	0.005281	7.70	0.008
Error	49	0.033613	0.000686		
Total	52	0.356252			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0261913	90.56%	89.99%	89.02%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.245	0.274	0.90	0.374	
Difficulty	0.2886	0.0289	9.99	0.000	1.31
Log Minerev	0.4199	0.0444	9.47	0.000	1.63
Log GoogleSearch	0.1025	0.0370	2.77	0.008	1.29

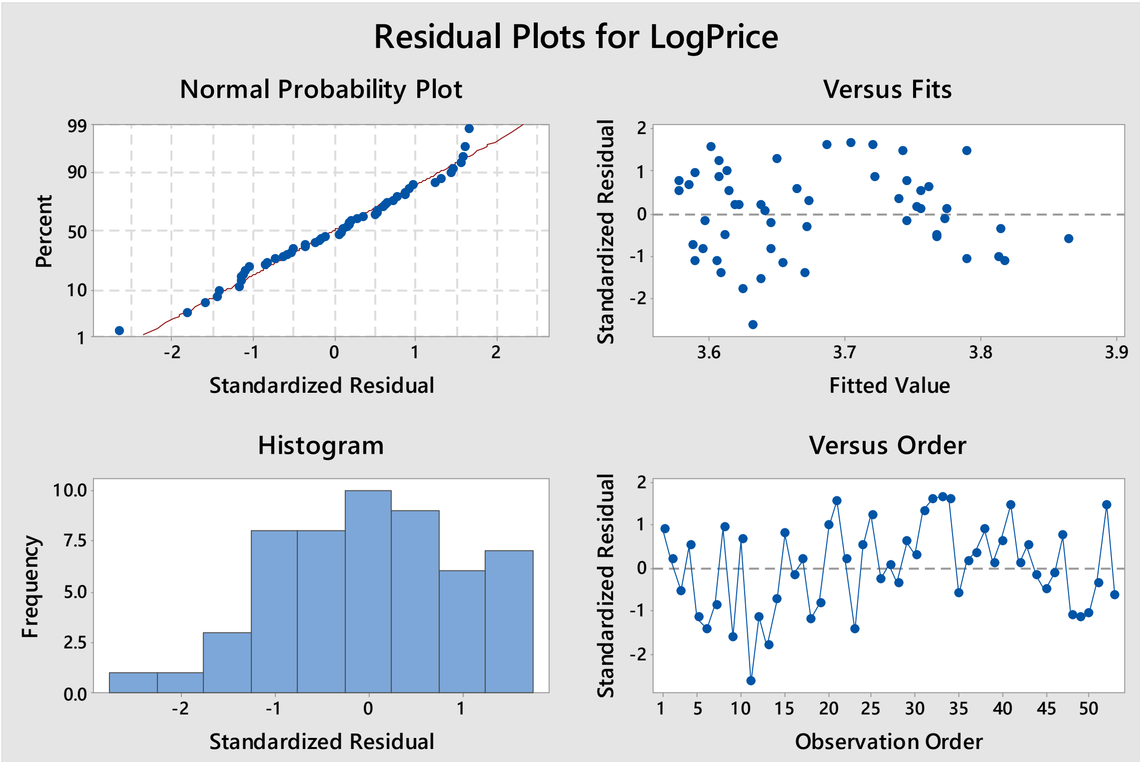
Regression Equation

LogPrice = 0.245 + 0.2886 Difficulty + 0.4199 Log Minerev + 0.1025 Log GoogleSearch

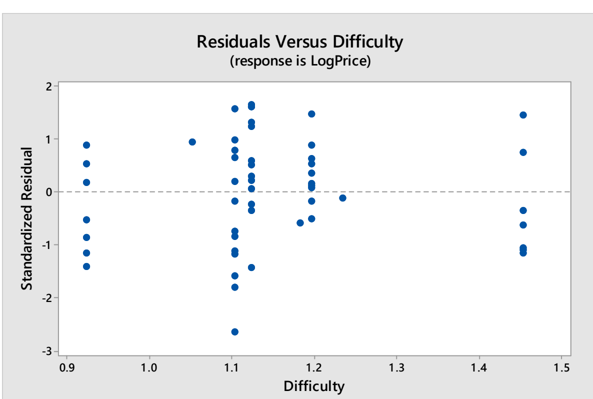
Durbin-Watson Statistic

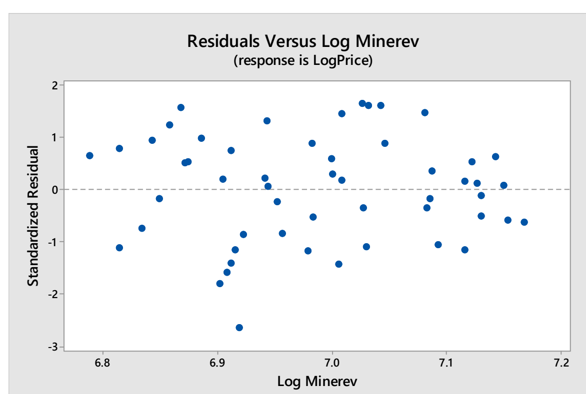
Durbin-Watson Statistic = 1.49785

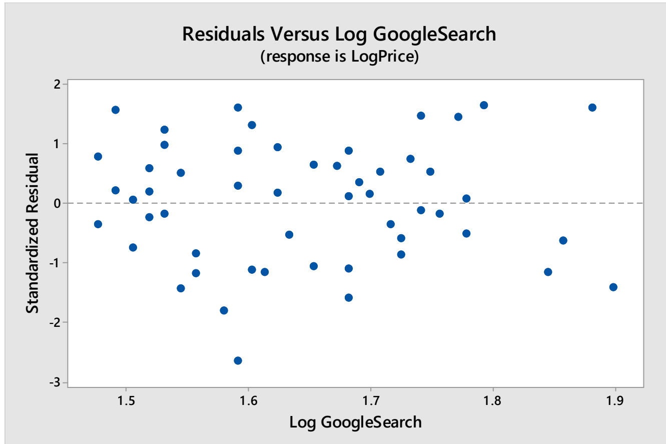
The regression model seems like a little improvement on the one in which the outliers and the leverage points were not handled. It shows a high R^2 of 90.56%, an error of 0.026, and high F-statistic. The adjusted error term of the model is 0.0261913 * sqrt((53-3-1)/(53-5-1)) = 0.027. This means that we could make a rough prediction of the Logged Market Price 95% of the time in the interval $10^{-2(0.027)}$ of the Market Price to $10^{2(0.027)}$ of the Market Price. This gives us an interval of 88% of the Market Price to 1.13 times the Market Price. The regression equation indicates that a 1-unit increase in the Difficulty is associated with $10^{0.2886}$ = 94.4% increase in the market price of bitcoin. A 1% increase in the Miners' Revenue corresponds to 0.42% increase in the market price holding everything else fixed. And a 1% increase in the Number of Google Searches corresponds to 0.1025% increase in the market price holding all else fixed. It's surprising to see a very large percentage change in the logged Market Price associated with the Difficulty predictor. I suspect that I need to take more properties of Bitcoin's Difficulty into account to better understand how it changes over time. It might be the case that the predictor variable difficulty has a special relation with the response variable, which demands further investigation.



Looking at the residual plots, we can observe normality in our data points. The residuals versus fits plot, however, did not change much even after correcting for the outliers. There is still a cyclic pattern present. We also need to take into account the fact that our dataset is small with 53 observations and in order to further correct for constant variance assumption, we should consider getting more data. Other than that, there does not appear to be any other major issue with our regression assumptions. Let's also take a look at the residuals versus each predictor.







The residuals versus each predictor plots all show constant variance. In the residuals versus Log GoogleSearch plot, there is a weak cyclic relation, but again we should be mindful of small sample size.

In conclusion, we can see that the predictor variables adding most to the predictive power of our model are Number of Google Searches of Bitcoin, Miners' Revenue, and Difficulty of Bitcoin Mining. We notice that Number of Google Searches gives us a rough estimate of the demand for Bitcoin. While the other two predictor variables are concerned with the supply of Bitcoin. This does support our initial hypothesis that Bitcoin's demand and supply play a vital role in estimating its value. We also noticed another interesting property of Bitcoin that warrants further investigation - its response to news and speculations. The outliers we identified, all corresponding to drastic rise or fall in either the Market Price or the Miners' Revenue, were followed by a major breaking news in some major media outlet. We observed the trend that people react to such news and their reaction also models the value of Bitcoin. So in the future, I would try to incorporate a variable that corresponds to news articles on Bitcoins. Furthermore, in my model, I have only used a few predictor variables out of the many potential variables that might impact the value of Bitcoin, so my model could be further improved by taking into account other variables as well. Other than that, I would also like to obtain more data and increase the time lags such that I work on weekly or monthly data instead of day-to-day data. This is because, from the charts spanning over a couple of months, I see a weak cyclic pattern in the fluctuation of Bitcoin value. This would also require us to test for autocorrelation again.

At the beginning of this analysis, I quoted Professor Milton Friedman stressing upon the need to reduce the role of government over money by introducing a reliable form of e-cash. I believe that the study of Bitcoin and its value would assist us in identifying the loopholes in Bitcoin, which would further result in major improvements in the technology backing it. A thorough study of Bitcoin's behavior would not only achieve that but will also help the potential investors invest smartly into Bitcoins. This report was just a starter to better understand the nature of Bitcoin and requires much more sophisticated data analyses techniques to further explore the cryptic nature of this cryptocurrency.

Works Cited

"'A Real Bubble': Billionaire Warren Buffett Doubles Down on Bitcoin Doubt." *CoinDesk*, 28 Oct. 2017, www.coindesk.com/real-bubble-billionaire-warren-buffett-doubles-bitcoin-doubt/

"Bitcoin Charts & Graphs - Blockchain." *Blockchain.info*, blockchain.info/charts.

Williams-Grut, Oscar. "Bitcoin's wild volatility continues with $500 swing." *Business Insider*, Business Insider, 15 Sept. 2017, www.businessinsider.com/bitcoin-ethereum-price-september-15-china-2017-9