ECON M 524 – Financial Econometrics
FINAL PROJECT

BITCOIN CLOSE PRICE PREDICTION

Introduction:

Bitcoin (BTC) is a cryptocurrency, a virtual currency designed to act as money and a form of payment outside the control of any one person, group, or entity, thus removing the need for third-party involvement in financial transactions.
Like all forms of money, Bitcoin is a brand-new digital currency that can be stored, exchanged, and used to make payments. Bitcoin's decentralized architecture and opt-in model are what distinguish it from national currencies like the US Dollar, the Euro, and the Japanese Yen. Why does that matter?

With centralized "fiat money" (literally, money by decree), central banks issue currency, and citizens are compelled to spend it. All transactions, except for cash (which is getting less and less common), go through middlemen like banks and payment gateways.
In contrast, Bitcoin is an opt-in currency that is governed by user consensus or will. It consists of an expanding network of users who willingly accept the Bitcoin protocol's guidelines. They employ decentralized infrastructure to carry out peer-to-peer transactions and hold value apart from any enterprise, government, or financial institution. When using Bitcoin, there is no requirement to request authorization and no chance of getting disconnected from the network.

Motivation:

Since technical analysis may allow investors to spot market trends and forecast the future price movements of an asset, traders must read cryptocurrency patterns to find the best opportunities in the market.

To understand how the supply and demand of a particular asset affect future price fluctuations, technical analysis refers to examining statistical trends gathered through time. Investors who want to know when positive and bearish moves will end can benefit from reading charts of the cryptocurrency market.

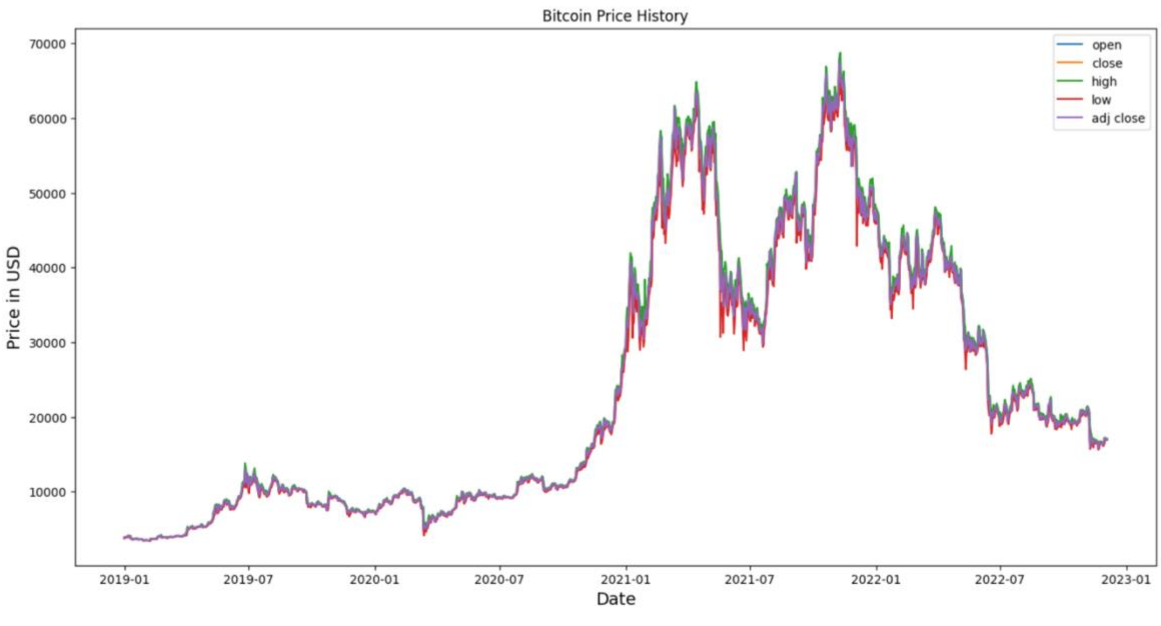
Data Description:

The data is collected from yahoo finance. In addition to stock quotes, news articles, financial reports, and unique material, it offers financial news, data, and opinion.

- Date: Dates from 1 January 2019 to 3 December 2022
- Open: The opening price of bitcoin on the date
- High: Maximum price of bitcoin
- Low: the minimum price of bitcoin
- Close: The closing price of bitcoin on the Date
- Adj Close: Adjusted closing price
- Volume: Cumulative sum of crypto being bought and sold on the market

Method:

By observing the trend of bitcoin history we can see that there is not much difference between open, close, low, high, and Adj close prices over time.



Candlestick graph:

Traders utilize candlestick charts to predict potential price movement based on historical trends.
When trading, candlesticks are helpful since they display four price points (open, close, high, and low) over the specified period.

The same price data displayed in candlestick charts serve as the foundation for numerous algorithms.
Trading decisions are frequently influenced by emotions, which are visible in candlestick charts.

It is very similar to the line graph plot over time indicating the market's open, high, low, and close prices for the day.

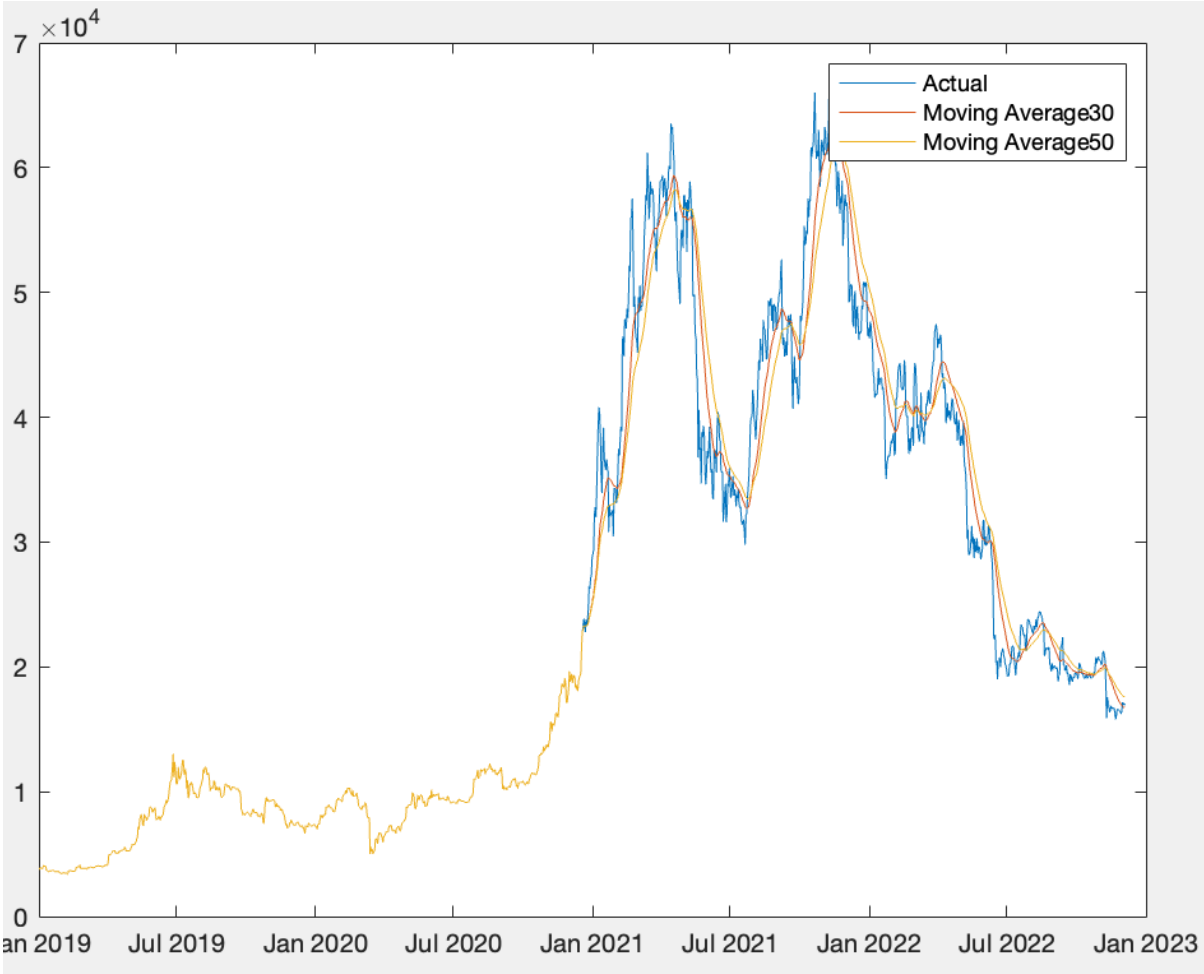
Moving Average (MA):

The moving average is a statistical method used for forecasting long-term trends. The technique represents taking an average of a set of numbers in a given range while moving the range. A moving average is a stock indicator commonly used in technical analysis, used to help smooth out price data by creating a constantly updated average price. A rising moving average indicates that the security is in an uptrend, while a declining moving average indicates a downtrend.

Autoregression (AR):

AR(p) is a regression model with lagged values of y, until the p-th time in the past, as predictors. Based on historical values, autoregressive models forecast future values. They are often used in technical analysis to predict upcoming stock price trends. The implicit assumption of autoregressive models is that the future will resemble the past.

Prediction Using Moving Average:



Fifty percent of the data is used to forecast the closing price of the bitcoin. Moving average predictions for a window size of 30 and 50 are predicted and compared. They both almost give the same results and seem to recognize all the patterns.

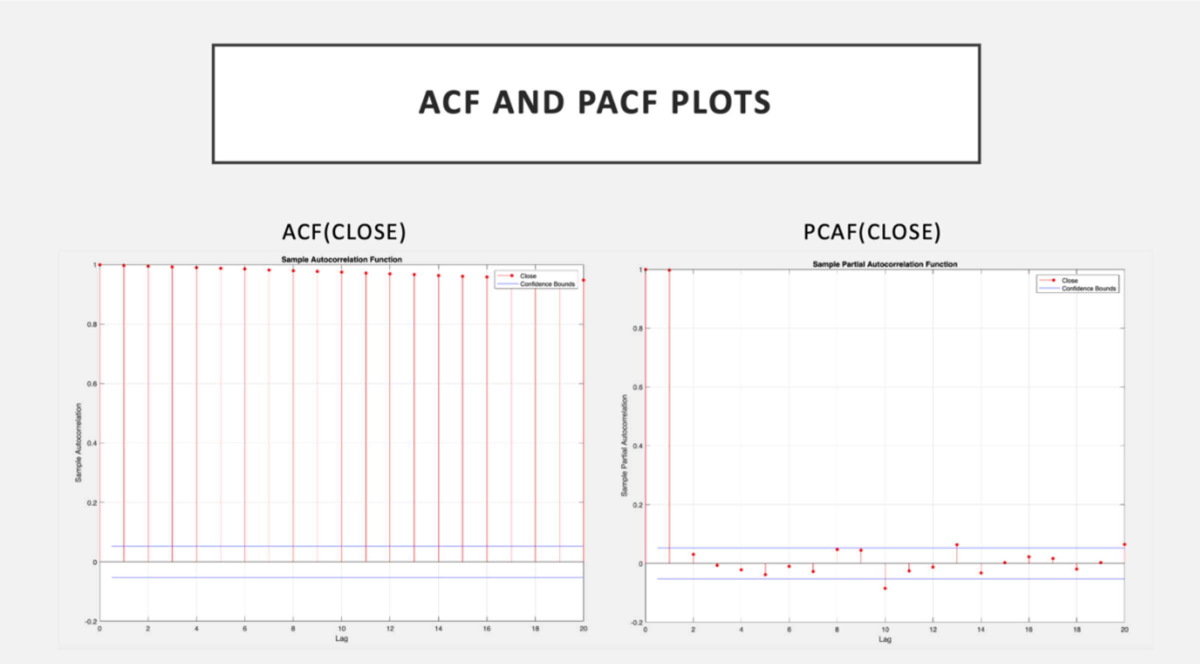
From the above plot, the blue line refers to the actual closing price of bitcoin over time. The red line is the prediction line made by the moving average with a window size of 30. The yellow line represents the prediction line made by the moving average of the window size 50.

ACF and PCAF:

To choose potential Auto Regressive Moving Average (ARMA) models for time series analysis and forecasting, one must understand the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the series to understand the AR and/or MA components' relative order. Although ACF and PACF don't directly influence the ARMA model's order, the plots can help with understanding the order and give a general notion of which model might fit the time-series data well.

ACF plots are bar graphs that display the coefficients of correlation between a time series and its lag values. Simply put, ACF explains the relationship between the present value of a particular time series and its past values (1-unit past, 2-unit past,..., n-unit past). The correlation coefficient is expressed on the y-axis of the ACF plot, while the number of lags is indicated on the x-axis. Assuming that the values of a time series at times t, t-1,..., t-n are y(t), y(t), y(t-1),..., y(t-n), the lag-1 value represents the correlation coefficient between y(t) and y(t), lag-2 represents the correlation coefficient between y(t) and y(t), and so on.

The partial autocorrelation function, or PACF, explains how the series and lags are somewhat correlated. Simple linear regression can be used to predict y(t) given y(t-1), y(t-2), and y(t-3) to explain PACF [2]. We use PACF to connect the "parts" of y(t) and y(t-3) that y(t-1) and y (t-2).

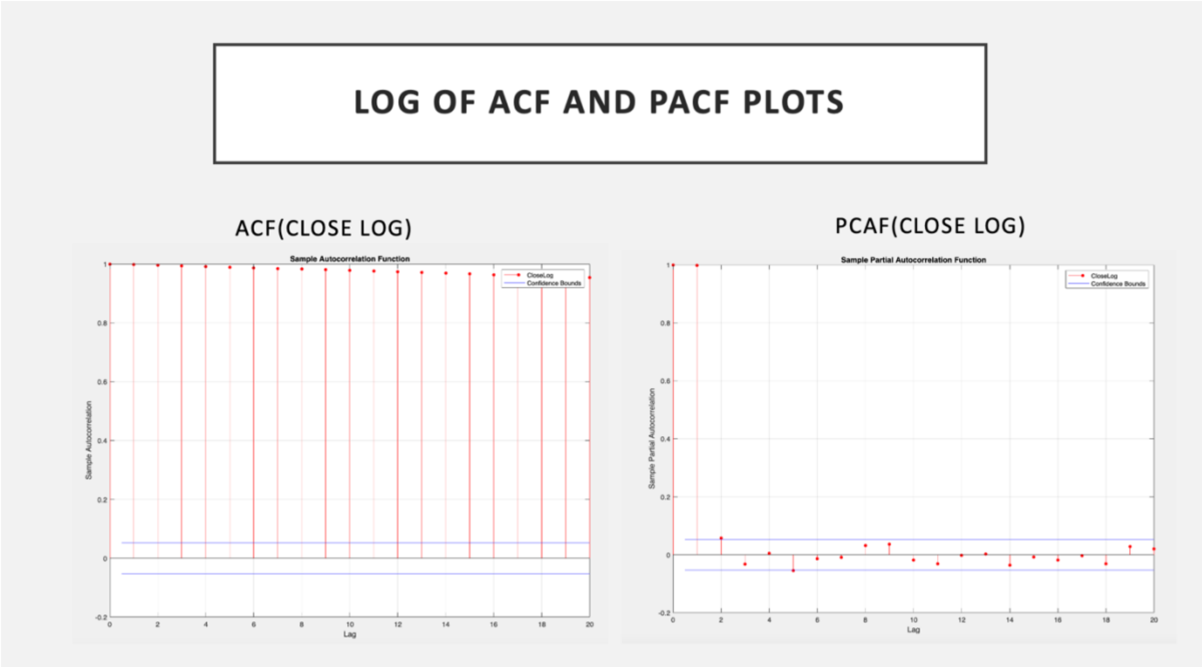


From the above plots, we can choose the order for Auto regression and moving average. But first when the KPSS test was performed the null hypothesis test was rejected stating that the close data is not stationary.

KPSS Test:

The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test determines if a time series is non-stationary because of a unit root or stationary around a mean or linear trend. When statistical characteristics, such as the mean and variance, remain constant across time, the time series is said to be stationary.

Therefore, the log of the closing price is taken to plot the ACF and PCAF plots. This will help in choosing a better model for the time series.



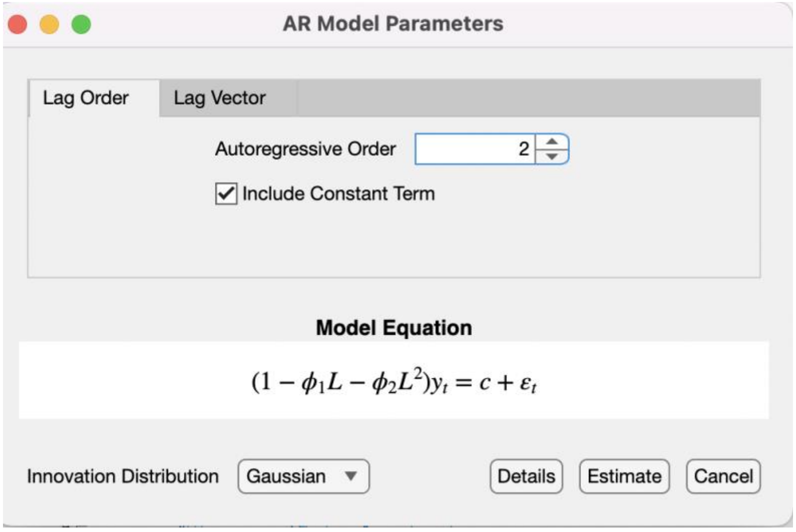
At times, only AR terms or only MA terms are sufficient to model the process. Firstly, we observe the ACF plot and see that there is a gradual decrease in the plot concerning the lags, to define the moving average. In the case of the gradual decrease in the ACF plot, there is no need of taking a moving average in the model.
On the other hand, PACF should have a sharp drop after p significant lags to define the AR of the model. In the above plot, it is clear that the PACF cuts immediately at two lags and then at 3 lags. In such cases, we could consider both AR (2) and AR(3) models which might be appropriate for the time series from the graphs for this data.

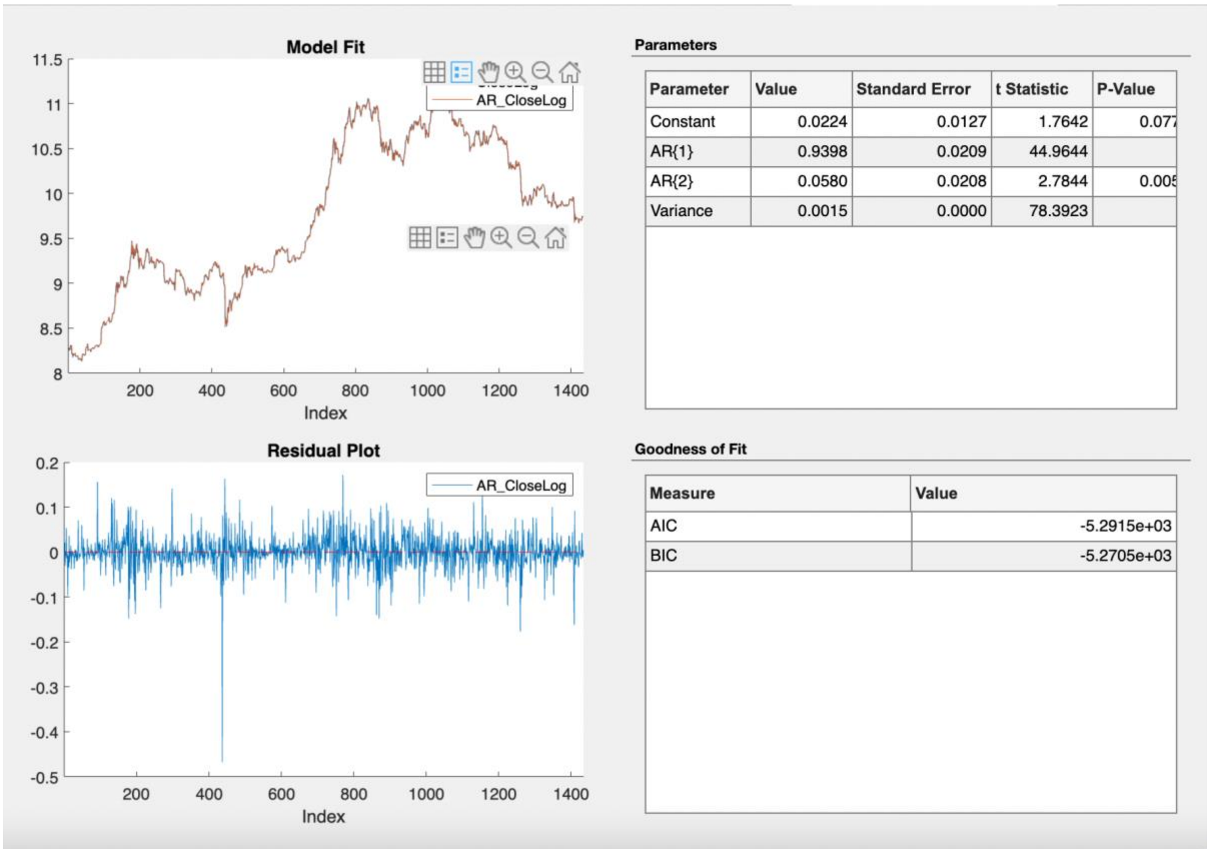
Analyzing the time series model using the econometric modeler:

An interactive tool for examining univariate or multivariate time series data is the Econometric Modeler app. The software is highly suited for performing statistical specification and model identification tests, fitting models to data, and iterating between these tasks, as well as for viewing and altering data. Once you are happy with a model, you can export it to the MATLAB® Workspace for further study or to predict future responses. A session can also be used to generate code or a report.

Implementing the AR(2) model to the data using the econometric modeler:

Select the AR model and enter the number of lags as 2 for AR(2) model.





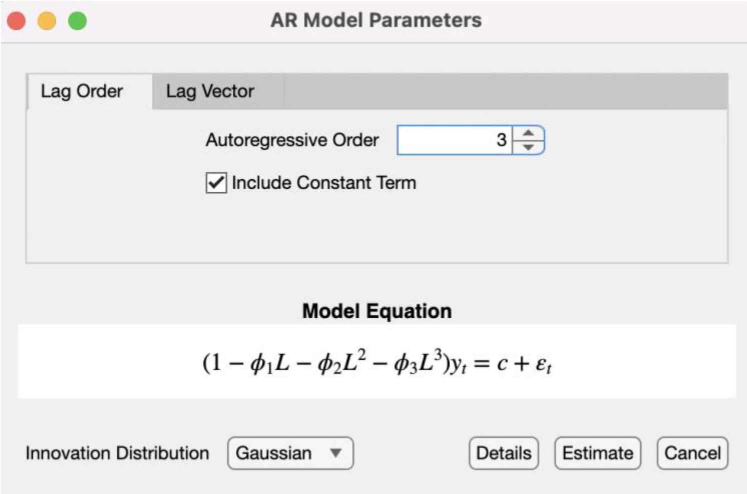
Parameter	Value	Standard Error	t Statistic	P-Value
Constant	0.0224	0.0127	1.7642	0.077
AR{1}	0.9398	0.0209	44.9644	
AR{2}	0.0580	0.0208	2.7844	0.005
Variance	0.0015	0.0000	78.3923	

Measure	Value
AIC	-5.2915e+03
BIC	-5.2705e+03

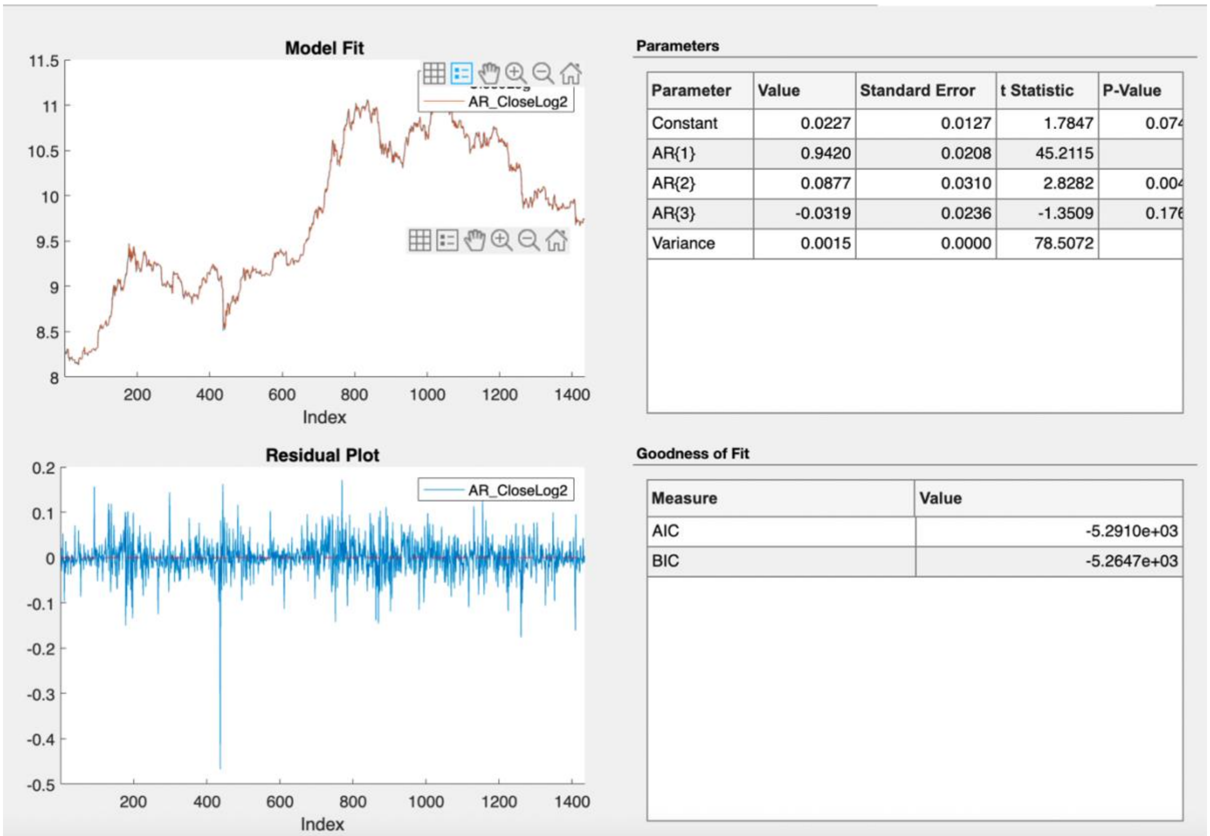
The above figure gives the results of the AR(2) model which is fit to the data.

Implementing the AR(3) model to the data using the econometric modeler:

Inserting the lags for the autoregressive model.



Results for the model fit for AR(3) :



AIC:

The Akaike information criterion (AIC) is a measure of the relative quality of statistical models for a given set of data and a predictor of prediction error. AIC calculates the quality of each model concerning the other models given a set of models for the data. As a result, AIC offers a model selection method.

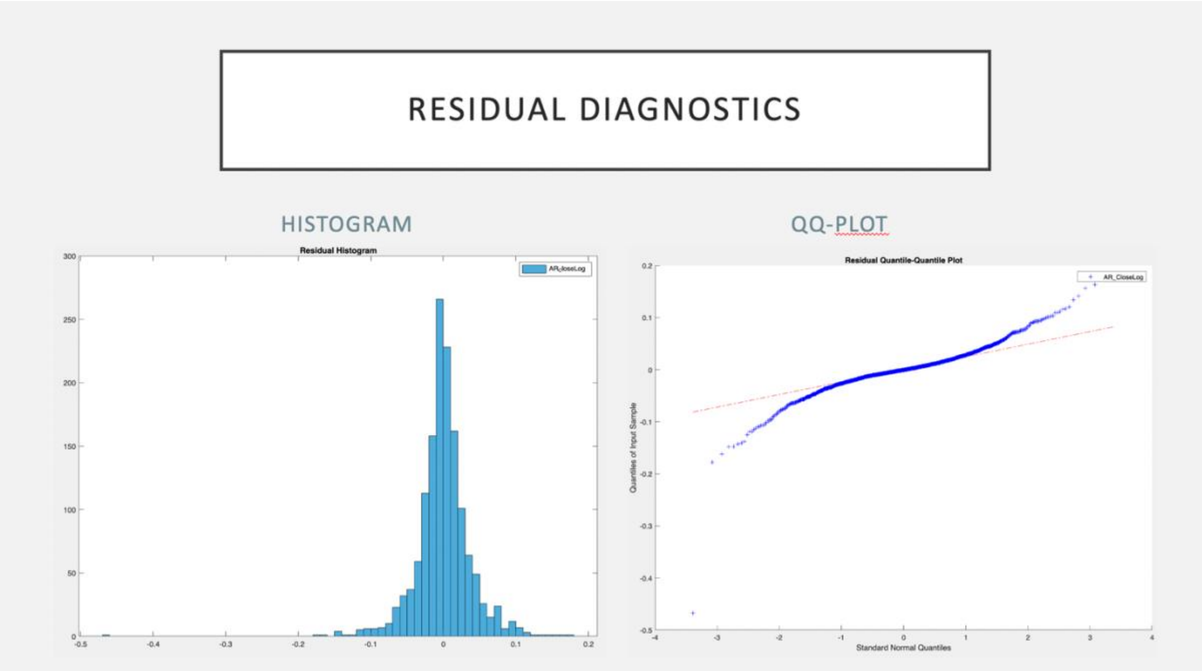
BIC:

The Bayesian Information Criteria (BIC) calculates the probability that a model will correctly predict. No expressly "good" BIC value exists. It is necessary to compare BIC values. The model with the lowest BIC value is the most appropriate one for the data.

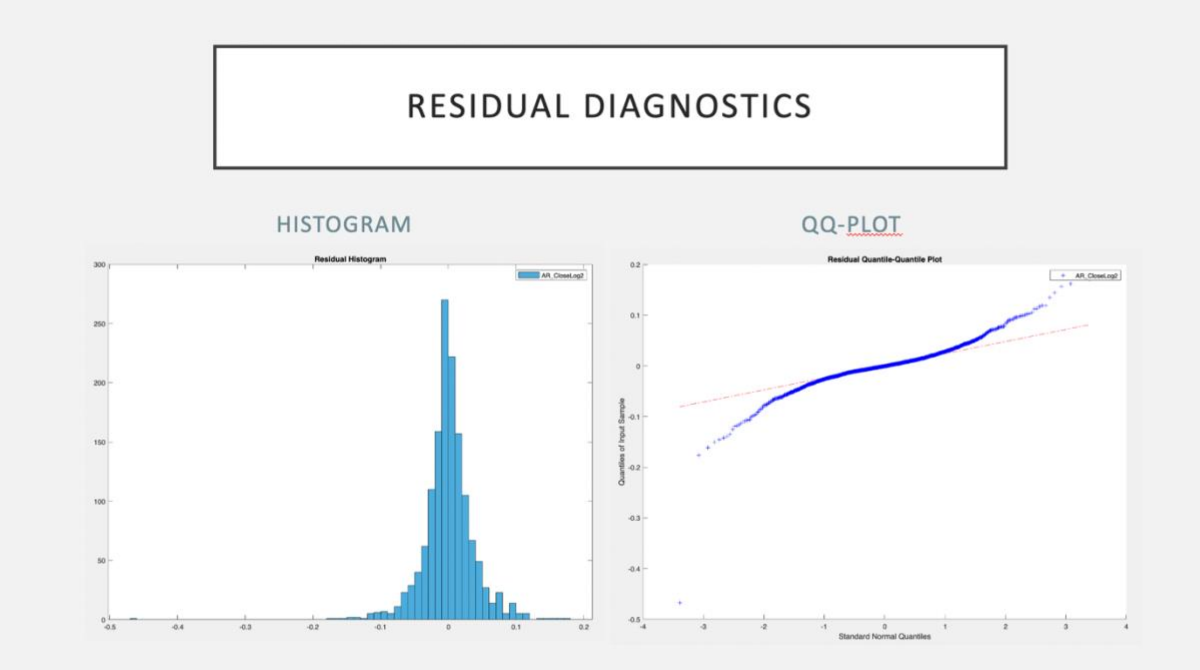
AIC penalizes models with additional parameters, and lower AIC scores are better. Therefore, the model with fewer parameters will have a lower AIC score and will be a better match if the two models explain the same amount of variation.

We can infer that the AR(2) and AR(3) models resemble a minute difference in the AIC and BIC values and the fit to the actual data of the bitcoin log of the closing price.

Another facility the econometric modeler provides is that to run the residual diagnostics. Exploring the diagnostics the results obtained were the following



The residual diagnostics for AR(2) model



The residual diagnostics for AR(3) model

Conclusion:

Moving averages with Windows 30,50 are almost similar. From the significant threshold of the ACF and PCAF graphs, the AR models are implemented without moving the average. AR(2) and AR(3) models resemble a minute difference in the AIC and BIC values.

Code:

```
bitcoin_data = readtable('BTC-USD.csv');
open = bitcoin_data.Open;
close = bitcoin_data.Close;
high= bitcoin_data.High;
adj_close=bitcoin_data.AdjClose;
low=bitcoin_data.Low;
date=datetime(bitcoin_data.Date);
vol=bitcoin_data.Volume;
f_start=0.5;
r=open;
n=length(r);
n1=fix(n*f_start);  %estimation sample
n2=n-n1;

%For h = 1
%aggregate dependent variable (containing true values)
h = 1;
y=zeros(n-h+1,1);
for i=1:(n-h+1);
   s=0;
   for j=0:(h-1);s=s+r(i+j);end;
   y(i)=s;
end;
y_true=y((n1+1):(n-h+1));
```

```
MSE_0=mean((0-y_true).^2);



%Moving Average Method
y_MA30=zeros(n2-h+1,1);
type = 'linear';
y_MA30 = movavg(y((n1+1):(n-h+1)),type,30);
MSE30=mean((y_MA30-y_true).^2)

y_MA50=zeros(n2-h+1,1);
type = 'linear';
y_MA50 = movavg(y((n1+1):(n-h+1)),type,50);
MSE50=mean((y_MA50-y_true).^2)


%These are for plotting purpose
yplot_true=[y(1:n1);y_true];
yplot_MA30=[y(1:n1);y_MA30];
yplot_MA50=[y(1:n1);y_MA50];

plot(date,yplot_true)
hold on
plot(date,yplot_MA30)
plot(date,yplot_MA50)
hold off
legend('Actual','Moving Average30','Moving Average50');
autocorr(bitcoin_data)
parcorr(bitcoin_data)
```