Ethereum price Prediction

206001F - MM. AABIDH

Table of Contents

Univariate Time series model
Introduction

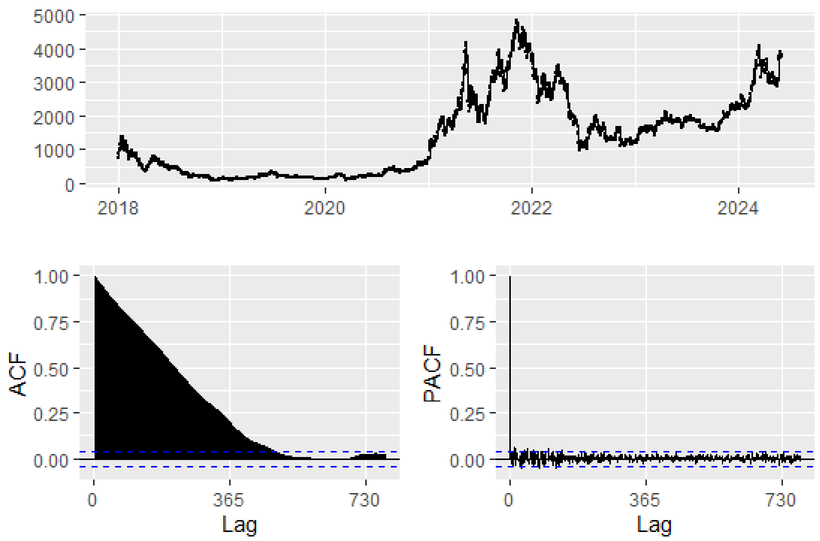
Since its launch in 2015, Ethereum, the second-largest cryptocurrency by market value, has experienced unusually high price volatility. Ethereum has grown in popularity over time as a platform for smart contracts and decentralized apps in addition to being a popular digital currency. Trading, investors, and other blockchain ecosystem participants need to be aware of and able to forecast Ethereum's price fluctuations.

Ethereum has seen a range of market dynamics between January 1, 2018, and June 1, 2024, including bull runs, bear markets, changes in regulations, technological breakthroughs, and adoption trends. Analyzing Ethereum's historical price data over this period might assist in predicting future patterns and offer insightful information about the pricing behavior of the cryptocurrency.

Based on historical data, time series analysis techniques like the SARIMA model can be used to model the price movements of Ethereum. We can try to estimate Ethereum's price trajectory within the given timeframe by using past Ethereum price data and predictive algorithms like SARIMA. This analysis aims to provide stakeholders and investors with useful information as they navigate the ever-changing cryptocurrency market.

EDA

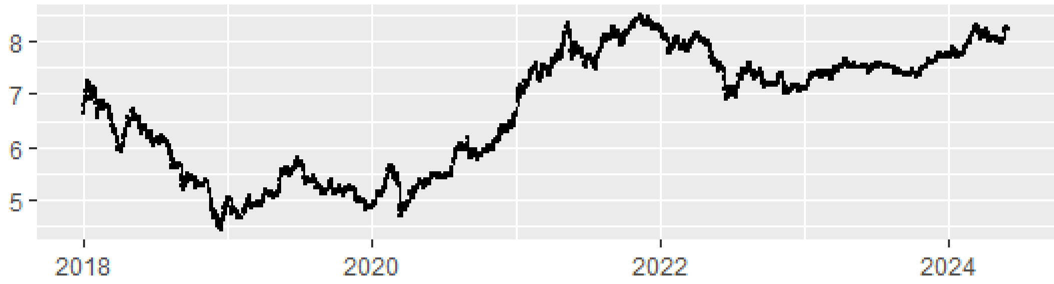
Firstly, let's illustrate the historical trends of Ethereum's value by charting its price fluctuation s over time.



```
[1] 0.0865862
```

Next, we'll assess the normality of the Ethereum price data. To do this, we'll apply the Box-Cox transformation, which stabilizes variance and reduces skewness in time-series data.

After applying the transformation Ethereum price graph



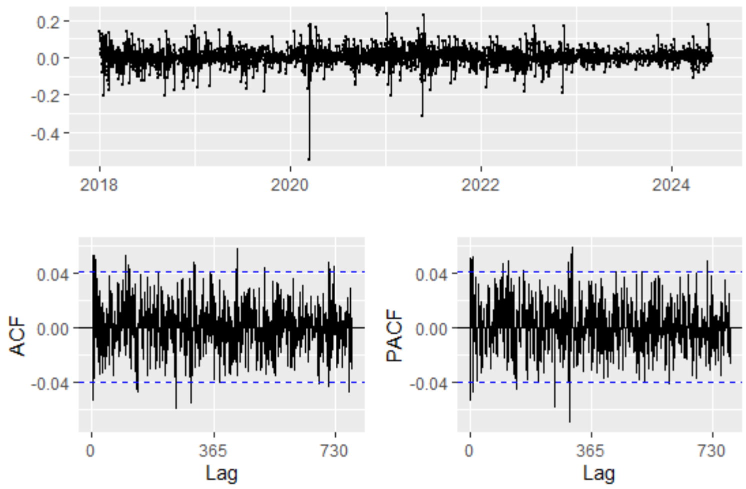
Checking stationarity

To determine non-stationarity, we use augmented dickey fuller and Phillips-perron.

```
        Phillips-Perron Unit Root Test

data:  eth_bc
Dickey-Fuller = -2.1077, Truncation lag parameter = 8, p-value = 0.5327
Augmented Dickey-Fuller Test

data:  eth_bc
Dickey-Fuller = -2.762, Lag order = 13, p-value = 0.2557
alternative hypothesis: stationary
```

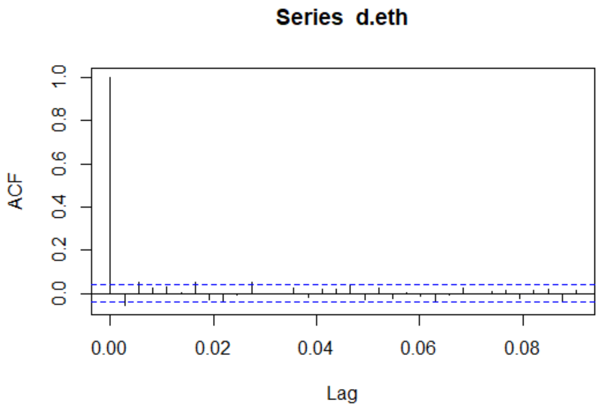
Since the process has been shown as non-stationary, we first difference the process Then we rechecked stationarity using the augmented Dickey fuller test and Phillips-perron we got the result like this.

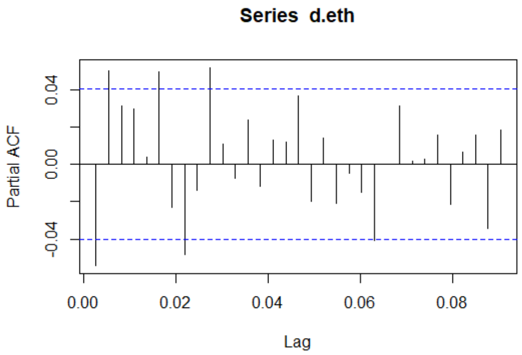
```
        Phillips-Perron Unit Root Test

data:  d.eth
Dickey-Fuller = -51.104, Truncation lag parameter = 8, p-value = 0.01
        Augmented Dickey-Fuller Test

data:  d.eth
Dickey-Fuller = -12.484, Lag order = 13, p-value = 0.01
alternative hypothesis: stationary
```



We successfully transformed our time series into a stationary process by applying first differencing. The above-augmented dickey fuller test, Phillips-perron, and the graph show the stationary process.

Checking ACF and PACF





According to ACF and PACF graph p should be equal to 2 and q should be equal to range from 1 or 2.

Model selection.

Utilizing the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), we evaluate a variety of combinations of 'p' (lag order for the AR part) and 'q' (lag order for the MA part) for a time series model. Subsequently, we compare the performance of these models using the Akaike Information Criterion (AIC), Akaike Information Criterion with Correction (AICc), and Bayesian Information Criterion (BIC) to determine the most suitable ARIMA model.

```
Series: eth_bc
ARIMA(2,1,0)

Coefficients:
          ar1     ar2
      -0.0512  0.0505
s.e.   0.0207  0.0207

sigma^2 = 0.002175:  log likelihood = 3858.81
AIC=-7711.62   AICc=-7711.61   BIC=-7694.34
> arima2  # has the lowest AIC, AICC and BIC
Series: eth_bc
ARIMA(2,1,1)

Coefficients:
         ar1     ar2      ma1
      0.5719  0.0865  -0.6259
s.e.  0.1752  0.0206   0.1754

sigma^2 = 0.002171:  log likelihood = 3861.17
AIC=-7714.33   AICc=-7714.32   BIC=-7691.3
> arima3
Series: eth_bc
ARIMA(1,1,2)

Coefficients:
         ar1      ma1     ma2
      0.6661  -0.7202  0.0869
```

```
s.e.  0.1621   0.1621  0.0207

sigma^2 = 0.002171:  log likelihood = 3861.19
AIC=-7714.38   AICc=-7714.37   BIC=-7691.35
> arima4
Series: eth_bc
ARIMA(2,1,1)

Coefficients:
         ar1     ar2      ma1
      0.5719  0.0865  -0.6259
s.e.  0.1752  0.0206   0.1754

sigma^2 = 0.002171:  log likelihood = 3861.17
AIC=-7714.33   AICc=-7714.32   BIC=-7691.3
> arima5
Series: eth_bc
ARIMA(2,1,2)

Coefficients:
         ar1     ar2      ma1     ma2
      0.5989  0.0460  -0.6531  0.0408
s.e.  0.2763  0.2936   0.2772  0.2946

sigma^2 = 0.002172:  log likelihood = 3861.19
AIC=-7712.38   AICc=-7712.36   BIC=-7683.59
```

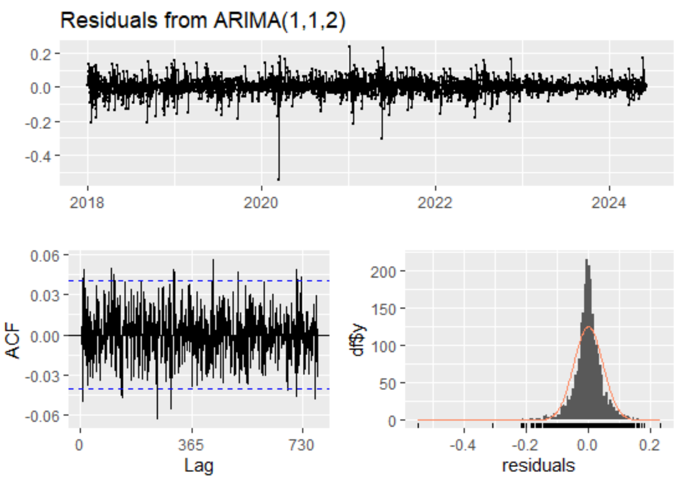
Checking auto Arima

```
arima_auto
Series: eth_bc
ARIMA(1,1,2)

Coefficients:
         ar1      ma1     ma2
      0.6661  -0.7202  0.0869
s.e.  0.1621   0.1621  0.0207

sigma^2 = 0.002171:  log likelihood = 3861.19
AIC=-7714.38   AICc=-7714.37   BIC=-7691.35
```

Significant outcomes were obtained using the ARIMA 3 model as well as the automatically chosen ARIMA model via the auto ARIMA function.

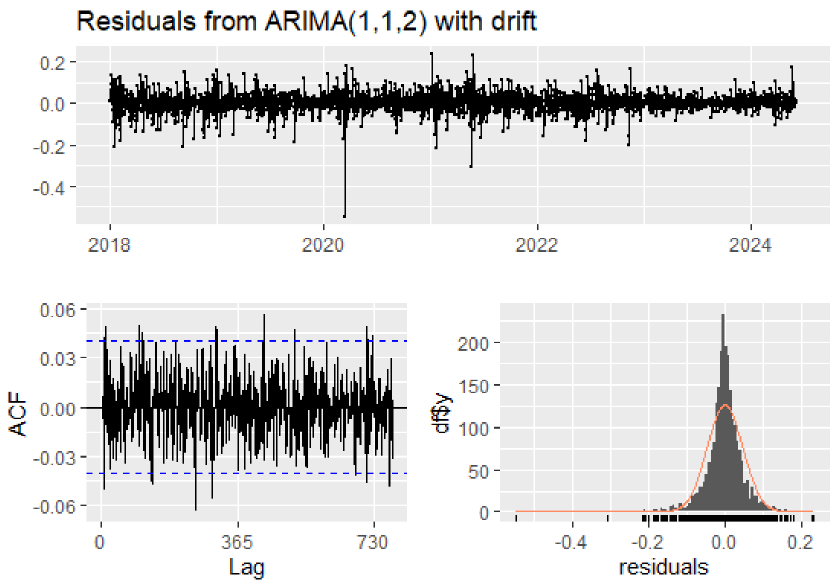
Upon selecting the ARIMA model, it was found that the errors were not serially correlated. The Ljung-Box test further confirmed this, indicating that the errors are, in fact, white noise.

```
> checkresiduals(arima3)

	Ljung-Box test

data:  Residuals from ARIMA(1,1,2)
Q* = 468.86, df = 466, p-value = 0.4541

Model df: 3.   Total lags used: 469
```



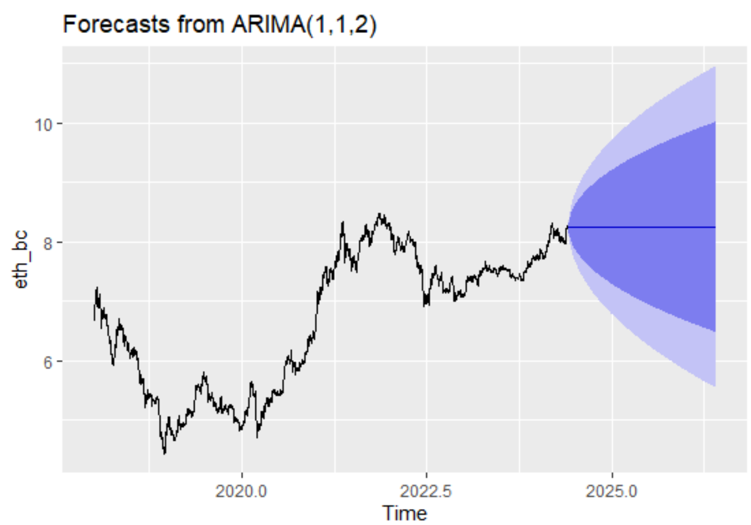
Next, we checked Arima with drift.

```
> arima6
Series: eth_bc
ARIMA(1,1,2) with drift

Coefficients:
         ar1      ma1     ma2   drift
      0.6651  -0.7195  0.0868  0.0007
s.e.  0.1594   0.1594  0.0207  0.0011

sigma^2 = 0.002172:  log likelihood = 3861.41
AIC=-7712.82   AICc=-7712.79   BIC=-7684.02
> checkresiduals(arima6)

	Ljung-Box test

data:  Residuals from ARIMA(1,1,2) with drift
Q* = 468.88, df = 466, p-value = 0.4538

Model df: 3.   Total lags used: 469
```



Upon applying the ARIMA model with drift, we proceeded to conduct the Ljung-Box test on the residuals of the ARIMA (1,1,2) model. The results of the test revealed yielding a p-value of 0.4538. These findings suggest that the errors in our model can be considered white noise.

The below line graph shows a forecast of the price of a stock using an auto plot. The forecast is based on an ARIMA (1,1,2) model and auto Arima since they got the same value.



Model Training and validation

The whole dataset was partitioned into two segments train and test.

```
> print(start(eth.train))
[1] 2018    1
> print(end(eth.train))
[1] 2024    1
> print(start(eth.test))
[1] 2024    2
> print(end(eth.test))
[1] 2024  154
```

Then we Check the accuracy of the above model

```
> accuracy(forecast(arima.train, h=120), eth.test)
                      ME      RMSE       MAE        MPE     MAPE       MASE         ACF1
Training set  -0.04500507  74.89154  39.60739 -0.2428612  3.21506 0.03761128 0.0008556777
Test set     677.05789979 850.71094 686.42934 20.1105788 20.52881 0.65183501 0.9732732483
             Theil's U
Training set        NA
Test set      7.422753
> accuracy(forecast(arima.train2, h=120), eth.test)
                     ME      RMSE      MAE         MPE      MAPE       MASE         ACF1
Training set   0.6831961  74.89486  39.6013 -0.06321808  3.201828 0.03760549 0.0006524283
Test set     721.2859522 897.18796 727.6572 21.50811385 21.793528 0.69098507 0.9740289038
             Theil's U
Training set        NA
Test set      7.850071
> accuracy(forecast(arima.train5, h=120), eth.test)
                      ME      RMSE       MAE        MPE      MAPE       MASE         ACF1
Training set  -0.02377479  74.52444  39.35962 -0.2469339  3.221456 0.03737599 -0.003670625
Test set     679.43646214 852.88968 688.50143 20.1892273 20.593994 0.65380267  0.973393675
             Theil's U
Training set        NA
Test set      7.442908
```

RMSE - When compared to Model 2(Arima 1,1,2), Models 1(Arima 1,1,2 with drift) and 3(Arima 2,1,2 with drift) have somewhat lower RMSE values, suggesting more accuracy in ETH price predictions.

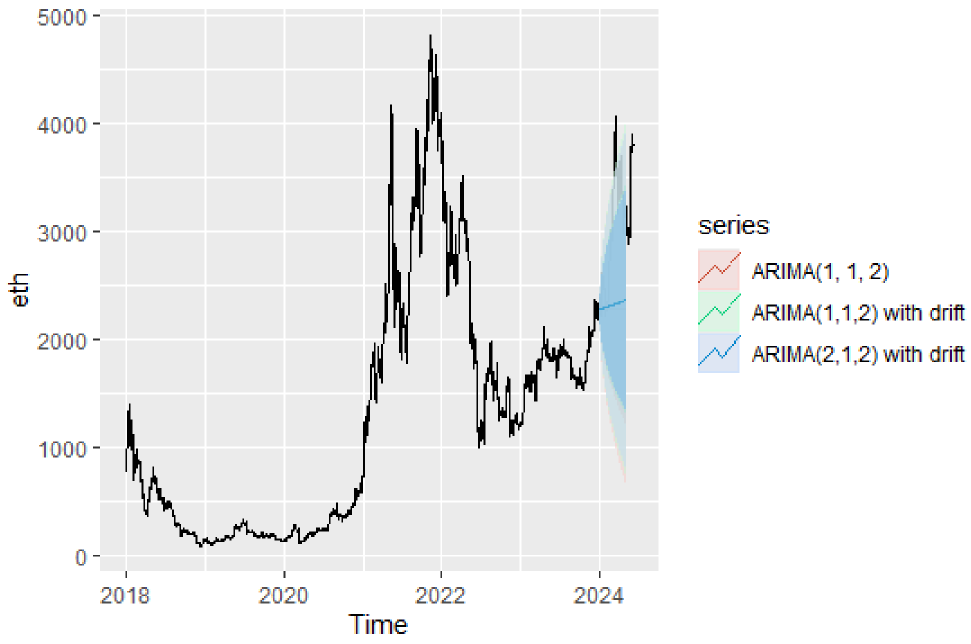
MAPE - In terms of percentage error, Models 1 and 3 appear to have a better match to the observed data, as indicated by their lower MAPE values.

Theil's U - Theil's U value for Model 1 is the lowest, suggesting greater overall forecast accuracy.

MASE - With a little lower MASE value than a naive forecast, Model 3 performs better in terms of prediction.

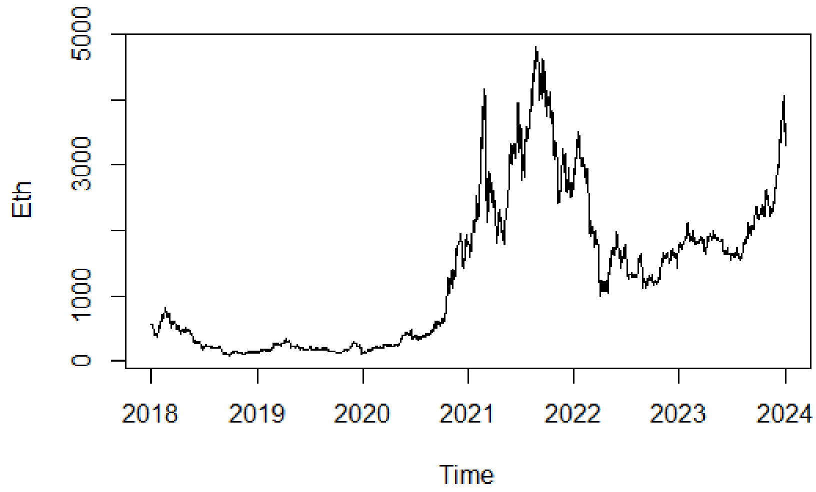
This information suggests that Model 1 performs somewhat better overall, with Model 3 following in a close second.

Forecasting



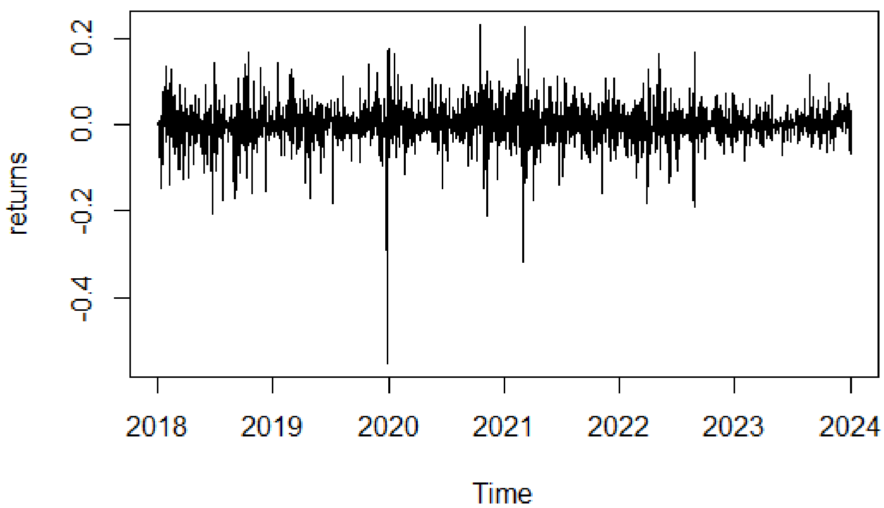
The model's ARIMA forecast for Ethereum prices is shown graphically. The predicted price ranges over the last 154 days are represented by the colored regions. larger areas denote higher uncertainty. It shows an overall increasing trend in Ethereum prices. Analysts can determine which ARIMA model provides the most dependable forecasts by looking at these various models, such as ARIMA (1, 1, 2) without drift, ARIMA (1, 1, 2) with drift, and ARIMA (2, 1, 2) with drift.

ARCH/GRACH model



The graph shows that the price of Ethereum has been volatile over the past few years. The graph does show slight signs of volatility clustering since periods of high volatility are followed by high volatility and low volatility periods are followed by low volatility. For example, the price of Ethereum increased rapidly in 2021, but then it decreased in 2022.

The logarithmic form of the dataset was then applied to determine the relative change in stock prices. To further capture series instability and guarantee series stationarity, the series was differenced once.

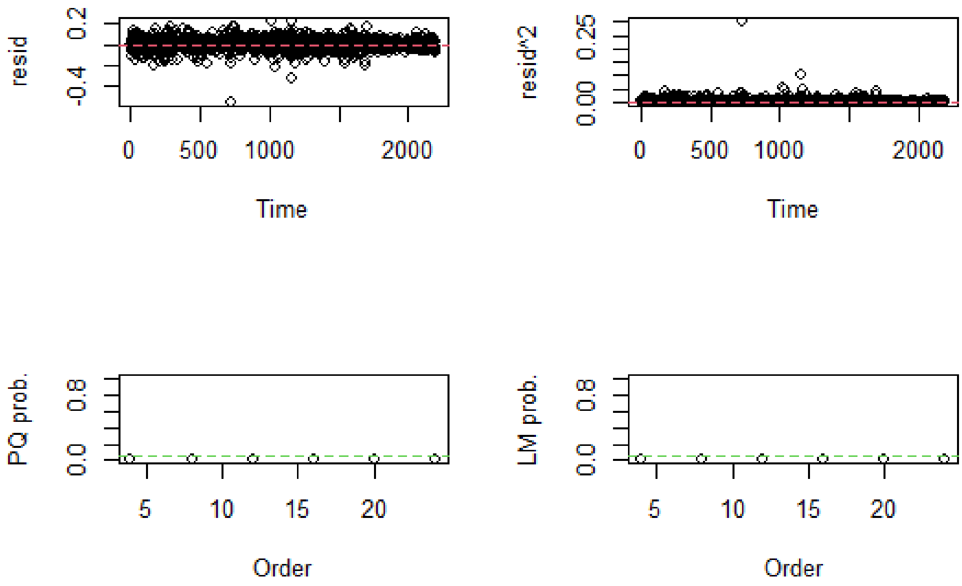


The graph indicates that the time series exhibits fluctuation across time. Volatility clustering, in which periods of low volatility are followed by low volatility and times of high volatility are followed by high volatility, may be responsible for this fluctuation.

Testing of ARCH

Testing of ARCH effects formally (order: AR (1), I(1), MA(2)) since, it got the best result in the Arima

```
ARCH heteroscedasticity test for residuals
alternative: heteroscedastic

Portmanteau-Q test:

     order   PQ  p.value
[1,]     4 43.1 9.66e-09
[2,]     8 65.9 3.18e-11
[3,]    12 70.3 2.85e-10
[4,]    16 73.3 2.62e-09
[5,]    20 75.3 2.44e-08
[6,]    24 76.4 2.24e-07
Lagrange-Multiplier test:
     order   LM p.value
[1,]     4 6978       0
[2,]     8 2646       0
[3,]    12 1755       0
[4,]    16 1314       0
[5,]    20 1029       0
[6,]    24  852       0
```



Less than 0.001 indicates that every PQ test p-value is extremely significant. This suggests that the null hypothesis that there are no ARCH effects at any of the lag orders examined can be rejected. That means the residuals of your ARIMA model clearly show the presence of ARCH effects.

Additionally, the LM test's p-values are very significant in this instance, which means its less than 0.001. This is consistent with the findings of the Portmanteau-Q test, indicating compelling proof of ARCH effects in the residuals of Ethereum price return data.

According to both tests, the residuals continuously display fluctuating variance over time, which is a feature of the ARCH effect, in which the variance at any one time is dependent on errors made in the past.

Fitting an ARCH (1) model

```
Title:
 GARCH Modelling

Call:
 garchFit(formula = ~garch(1, 0), data = returns, include.mean = FALSE,
    trace = FALSE)

Mean and Variance Equation:
 data ~ garch(1, 0)
<environment: 0x0000029cc99e4e90>
 [data = returns]

Conditional Distribution:
 norm

Coefficient(s):
    omega     alpha1
0.0019078  0.1010503

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
omega  0.0019078   0.0000702   27.176  < 2e-16 ***
alpha1 0.1010503   0.0279331    3.618 0.000297 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log Likelihood:
 3659.506    normalized:  1.669483
```

The distribution assumed for the error term under this model is normal.

Error analysis - The p values of both terms (omega and alpha) are significant indicating both terms are important in explaining the conditional variance of the Ethereum price returns.

```
Standardised Residuals Tests:
                                  Statistic      p-Value
 Jarque-Bera Test   R    Chi^2  1.791326e+04 0.000000e+00
 Shapiro-Wilk Test  R    W      9.116514e-01 0.000000e+00
 Ljung-Box Test     R    Q(10)  2.814213e+01 1.713024e-03
 Ljung-Box Test     R    Q(15)  3.316831e+01 4.448579e-03
 Ljung-Box Test     R    Q(20)  3.942077e+01 5.908136e-03
 Ljung-Box Test     R^2  Q(10)  5.151814e+01 1.401079e-07
 Ljung-Box Test     R^2  Q(15)  5.398450e+01 2.642524e-06
 Ljung-Box Test     R^2  Q(20)  5.860330e+01 1.168877e-05
 LM Arch Test       R    TR^2   4.822685e+01 2.853922e-06

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-3.337140 -3.331946 -3.337142 -3.335242
```

The Jarque-Bera test checks whether the residuals follow a normal distribution. A very low p-value indicates that the residuals do not follow a normal distribution. The Shapiro-Wilk test also assesses normality. A p-value of zero indicates strong evidence against the normality of the residuals. However, the normality of the residuals is not a strict requirement for ARCH models.

LM-ARCH test

(H0) - no ARCH effects

(H1) - there is an ARCH effect

The result of the LM Arch test indicates that there might still be some ARCH effects present in a model since its p-value is less than 5% significant.

Fitting an ARCH (2) model

Based on the LM Arch test results, the data set is set to a higher order model ARCH (2), assuming that the squared errors of the past two periods influence the current volatility.

```
Title:
 GARCH Modelling

Call:
 garchFit(formula = ~garch(2, 0), data = returns, include.mean = FALSE,
    trace = FALSE)

Mean and Variance Equation:
 data ~ garch(2, 0)
<environment: 0x0000028848ba0048>
 [data = returns]

Conditional Distribution:
 norm

Coefficient(s):
    omega     alpha1     alpha2
0.0016802  0.0987087  0.1253845

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
omega  1.680e-03   7.482e-05   22.457  < 2e-16 ***
alpha1 9.871e-02   2.873e-02    3.436 0.000591 ***
alpha2 1.254e-01   2.803e-02    4.474 7.69e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log Likelihood:
 3677.486    normalized:  1.677685

Description:
 Fri May 24 14:08:08 2024 by user: aabidh musthaq


Standardised Residuals Tests:
                                   Statistic      p-Value
 Jarque-Bera Test   R    Chi^2  2.426955e+04 0.0000000000
 Shapiro-Wilk Test  R    W      9.098226e-01 0.0000000000
 Ljung-Box Test     R    Q(10)  2.133639e+01 0.0188662507
 Ljung-Box Test     R    Q(15)  2.646395e+01 0.0334198896
 Ljung-Box Test     R    Q(20)  3.291651e+01 0.0344592920
 Ljung-Box Test     R^2  Q(10)  3.414456e+01 0.0001745675
 Ljung-Box Test     R^2  Q(15)  3.558954e+01 0.0020253181
 Ljung-Box Test     R^2  Q(20)  3.830258e+01 0.0081311185
 LM Arch Test       R    TR^2   3.363277e+01 0.0007705640

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-3.352633 -3.344842 -3.352637 -3.349786
```

all p-values are significant, suggesting that all three parameters play a vital role in explaining the conditional variance of the Ethereum price returns. Jarque-Bera and Shapiro-Wilk tests:

since the p-values are high it indicates that the residuals are not normally distributed. This is not a problem for ARCH models as the residuals are not autocorrelated.

The LM Arch test still with a less than 5% significant p-value suggests that there are ARCH effects in the residuals not captured. However, it is not as significant as the ARCH (1) model.

```
Information Criterion for arch1 Statistics:
      AIC       BIC       SIC      HQIC
-3.337140 -3.331946 -3.337142 -3.335242
Information Criterion for arch 2 Statistics:
      AIC       BIC       SIC      HQIC
-3.352633 -3.344842 -3.352637 -3.349786
```

Since arch2 has lower values for all information criteria (AIC, BIC, SIC, HQIC), it is the better model according to these metrics.

Fitting an ARCH (3) model

```

Title:
 GARCH Modelling

Call:
 garchFit(formula = ~garch(3, 0), data = returns, include.mean = FALSE,
    trace = FALSE)

Mean and Variance Equation:
 data ~ garch(3, 0)
<environment: 0x0000028847dc85b0>
 [data = returns]

Conditional Distribution:
 norm

Coefficient(s):
    omega     alpha1     alpha2     alpha3
0.0016037  0.0849170  0.1078636  0.0625603

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
omega  1.604e-03   7.396e-05   21.683  < 2e-16 ***
alpha1 8.492e-02   2.740e-02    3.099  0.00194 **
alpha2 1.079e-01   2.659e-02    4.057 4.98e-05 ***
alpha3 6.256e-02   1.949e-02    3.210  0.00133 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log Likelihood:
 3686.15    normalized:  1.681638

Description:
 Fri May 24 14:15:32 2024 by user: aabidh musthaq


Standardised Residuals Tests:
                                   Statistic    p-Value
 Jarque-Bera Test   R    Chi^2  2.896625e+04 0.00000000
 Shapiro-Wilk Test  R    W      9.068419e-01 0.00000000
 Ljung-Box Test     R    Q(10)  1.916899e+01 0.03816799
 Ljung-Box Test     R    Q(15)  2.390811e+01 0.06667054
 Ljung-Box Test     R    Q(20)  3.052223e+01 0.06182207
 Ljung-Box Test     R^2  Q(10)  2.034474e+01 0.02615556
 Ljung-Box Test     R^2  Q(15)  2.140620e+01 0.12434527
 Ljung-Box Test     R^2  Q(20)  2.390517e+01 0.24655907
 LM Arch Test       R    TR^2   2.085917e+01 0.05248111

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-3.359626 -3.349238 -3.359632 -3.355829
```

The same procedure is carried out for an ARCH (3) model. The coefficients of all term's omega, alpha1, alpha2, and alpha3 are significant indicating they are all important parameters in explaining the conditional variance of the Ethereum price returns. LM Arch tests' p-value is no longer significant at 5% suggesting that the ARCH (3) model captures all ARCH effects. Based on all these tests, the squared residuals from the ARCH (3) model seem to approximate white noise more closely than those from the ARCH (1) and ARCH (2) models. This indicates that the ARCH (3) model is more effective in capturing the volatility dynamics of the data.

GARCH models

```
Title:
 GARCH Modelling

Call:
 garchFit(formula = ~garch(1, 1), data = returns, include.mean = FALSE,
    trace = FALSE)

Mean and Variance Equation:
 data ~ garch(1, 1)
<environment: 0x0000028848a99010>
 [data = returns]

Conditional Distribution:
 norm

Coefficient(s):
     omega       alpha1        beta1
3.4772e-05   7.0763e-02   9.1756e-01

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
omega  3.477e-05   1.108e-05    3.140  0.00169 **
alpha1 7.076e-02   1.000e-02    7.075 1.49e-12 ***
beta1  9.176e-01   1.249e-02   73.445  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log Likelihood:
 3793.2    normalized:  1.730475


Standardised Residuals Tests:
                                   Statistic     p-Value
 Jarque-Bera Test   R    Chi^2  5187.4327436 0.000000000
 Shapiro-Wilk Test  R    W         0.9344623 0.000000000
 Ljung-Box Test     R    Q(10)    24.8213896 0.005694536
 Ljung-Box Test     R    Q(15)    29.1816105 0.015239841
 Ljung-Box Test     R    Q(20)    35.4276431 0.017941426
 Ljung-Box Test     R^2  Q(10)     8.9298301 0.538775693
 Ljung-Box Test     R^2  Q(15)    11.0226286 0.750988621
 Ljung-Box Test     R^2  Q(20)    14.6415449 0.796537948
 LM Arch Test       R    TR^2     10.4693205 0.574859168

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-3.458212 -3.450421 -3.458216 -3.455365
```

The GARCH (1,1) model can be written as,

conditional variance (t)= omega + alpha1 * squared residual of the previous period + beta1 * volatility from the previous period on the current volatility

GARCH (1,1) model is:

Omega: $3.4772\times10-5$

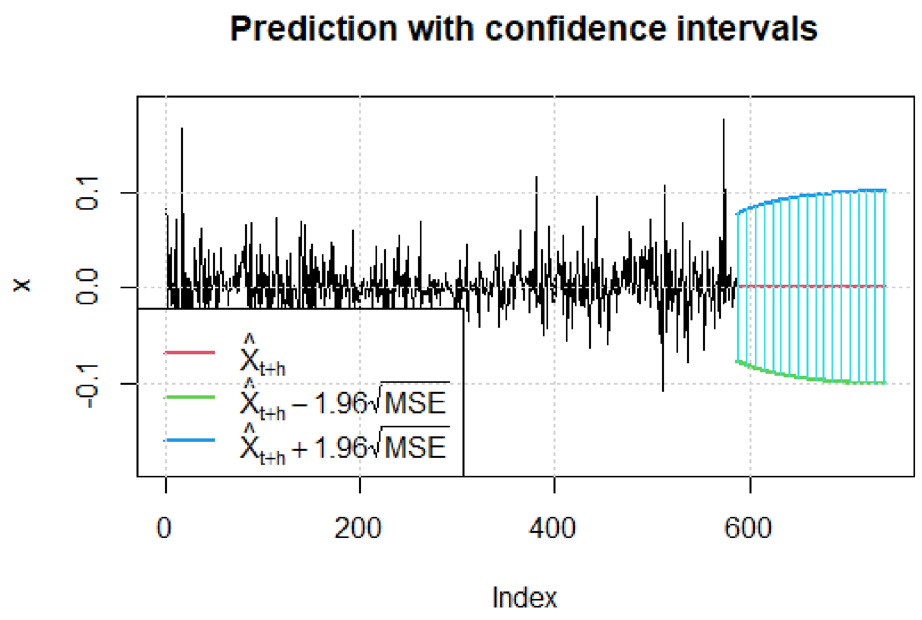
Alpha1: $0.070763\epsilon t-12$

Beta1: $0.91756\sigma t-12$

$\boldsymbol{\sigma t2=3.4772\times10-5+0.070763\epsilon t-12+0.91756\sigma t-12}$

All coefficients omega, alpha1, and beta1 are statistically significant which indicates the importance of explaining the conditional variance. According to the Jarque-Bera and Shapiro-Wilk tests, test statistics are extremely high, with a p-value of 0, suggesting non-normality. LM Arch test further indicates that there are no ARCH effects present in the residuals.

Predictions using GARCH (1,1) model



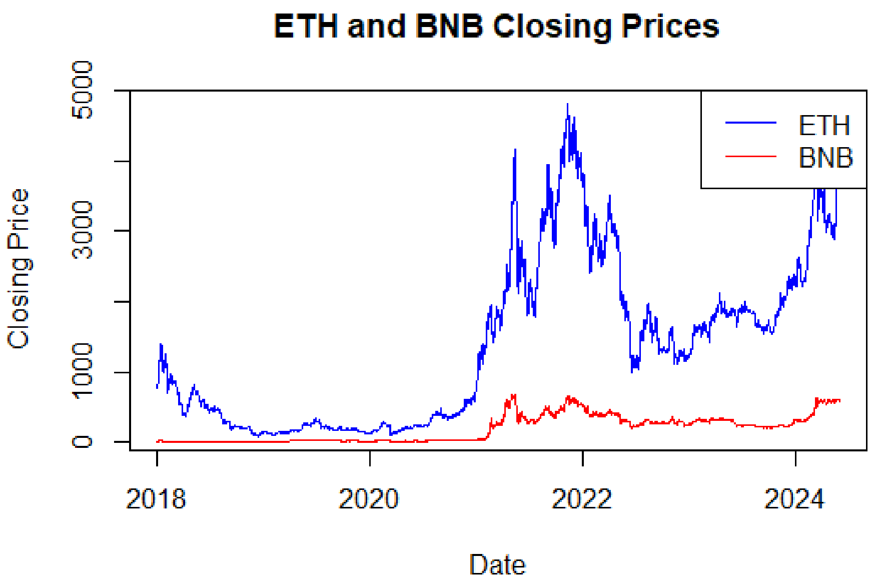
The graph indicates that within the next 154 days, the model predicts a gain in the price of Ethereum. When the prediction first starts, the confidence interval is larger and gets smaller over time. This may indicate that the model has greater confidence in its predictions on later days of the forecast.

Multivariate Time series model

Introduction

Multivariate time series models offer a robust framework for capturing the joint behavior of multiple time series. By incorporating BNB alongside ETH, we aim to enhance the predictive accuracy of our model, leveraging the potential correlations and cross-dependencies between these two prominent cryptocurrencies. First, we use the Engle-Granger and Johansen tests to find the best way to detect co-integration relationships. We use the Vector Autoregression (VAR) or Vector Error Correction Model (VECM) for our study depending on these findings. Because of its significant market presence and excellent connection with ETH, BNB has been selected to enhance forecasting performance overall and offer more in-depth insights into market trends (Binance, 2024).

EDA



Verma or VEC

Ethereum price

```
> PP.test(eth)

	Phillips-Perron Unit Root Test

data:  eth
Dickey-Fuller = -2.0959, Truncation lag parameter = 8, p-value = 0.5377

> adf.test(eth)

	Augmented Dickey-Fuller Test

data:  eth
Dickey-Fuller = -2.228, Lag order = 13, p-value = 0.4818
alternative hypothesis: stationary
```

BNB price

```
> PP.test(bnb)

        Phillips-Perron Unit Root Test

data:  bnb
Dickey-Fuller = -2.1617, Truncation lag parameter = 8, p-value = 0.5099

> adf.test(bnb)

        Augmented Dickey-Fuller Test

data:  bnb
Dickey-Fuller = -2.2499, Lag order = 13, p-value = 0.4725
alternative hypothesis: stationary
```

Since both p values are greater than 5% significance, we cannot reject null therefore processes are non-stationary

Engle-Granger cointegration model based on two equities baskets

```
Augmented Dickey-Fuller Test

data:  comb1$residuals
Dickey-Fuller = -3.8284, Lag order = 1, p-value = 0.01756
alternative hypothesis: stationary

> adf.test(comb2$residuals, k =1)

        Augmented Dickey-Fuller Test

data:  comb2$residuals
Dickey-Fuller = -3.9624, Lag order = 1, p-value = 0.01086
alternative hypothesis: stationary
```

Since the p-values for both tests are less than 0.05, we reject the null hypothesis of a unit root (non-stationarity). This means the residuals are stationary, indicating that the two-time series (ETH and BNB) are cointegrated.

```
Call:
lm(formula = d.eth ~ d.bnb + error.ecm1)

Residuals:
    Min      1Q  Median      3Q     Max 
-585.05  -12.05    0.20   13.59  480.32 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.16017    1.14394    0.14 0.888662    
d.bnb        4.47334    0.09346   47.86  < 2e-16 ***
error.ecm1  -0.07004    0.01951   -3.59 0.000338 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 55.36 on 2340 degrees of freedom
Multiple R-squared:  0.4954,	Adjusted R-squared:  0.495 
F-statistic:  1149 on 2 and 2340 DF,  p-value: < 2.2e-16
```

The coefficient −0.07004-0.07004−0.07004 indicates that 7.004% of the deviation from the equilibrium in the last period will be corrected in the next period.

Johanson test

```
> lagselect
$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
```

```
     7      7      2      7

$criteria
                  1            2            3            4            5            6
AIC(n)     13.03461     13.01986     13.01598     13.01513     13.00521     12.99414
HQ(n)      13.04001     13.02886     13.02858     13.03134     13.02501     13.01755
SC(n)      13.04943     13.04457     13.05056     13.05960     13.05955     13.05837
FPE(n) 457994.84169 451288.79974 449539.78549 449160.12683 444724.12737 439830.64804
                  7            8            9           10           11           12
AIC(n)     12.97515     12.97732     12.97786     12.97913     12.98103     12.98294
HQ(n)      13.00215     13.00792     13.01206     13.01694     13.02243     13.02794
SC(n)      13.04926     13.06130     13.07173     13.08289     13.09466     13.10645
FPE(n) 431554.76192 432491.11298 432727.07604 433278.45101 434100.44737 434931.46080
                 13           14           15
AIC(n)     12.98308     12.98517     12.98741
HQ(n)      13.03168     13.03737     13.04322
SC(n)      13.11647     13.12845     13.14057
FPE(n) 434989.48934 435902.37898 436881.34771
```

According to the AIC, HQ, and FPE, we can select 7 lags because it has the lowest among 15 lags.

There are two main types of Johansen test of cointegration

Lambda Trace test results

```
> vecm1 <- ca.jo(v1, type = "trace", K = 7, ecdet = "none", spec = "longrun")
> summary(vecm1)

######################
# Johansen-Procedure #
######################

Test type: trace statistic , with linear trend

Eigenvalues (lambda):
[1] 0.0081337104 0.0002247577

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 |  0.53  6.50  8.18 11.65
r = 0  | 19.61 15.66 17.95 23.52
```

H0 = no cointegrating vectors

H1 = at most one cointegrating vector

The findings of the Johansen Lambda Trace test show that there is proof of at least one cointegrating vector in the series. This indicates that the time series has a long-term equilibrium relationship, indicating that despite short-term aberrations, they move in the same direction over an extended period.

Lambda Max test results

```
> vecm2 <- ca.jo(v1, type = "eigen", K = 7, ecdet = "none", spec = "longrun")
> summary(vecm2)

######################
# Johansen-Procedure #
######################

Test type: maximal eigenvalue statistic (lambda max) , with linear trend

Eigenvalues (lambda):
[1] 0.0081337104 0.0002247577

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 |  0.53  6.50  8.18 11.65
r = 0  | 19.09 12.91 14.90 19.19
```

H0 = no cointegrating vectors

H1 = at most one cointegrating vector

The results of the Johansen maximal eigenvalue test show that the series appears to have at least one cointegrating vector. This indicates that the time series has a long-term equilibrium relationship, meaning that even with short-term aberrations, they move in tandem over time.

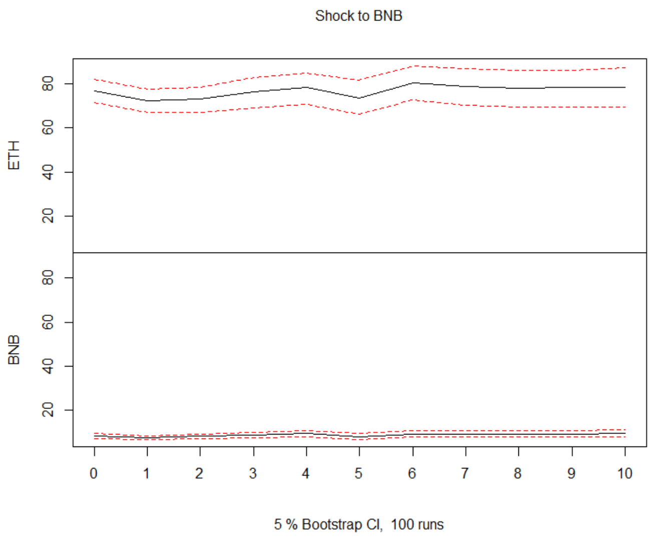
According to the above result in Engle-Granger cointegration, Lambda Trace test results, and Lambda Max test results suggest that we must use an error correction model in the multivariate time series.

Error Correction Model (ECM)

```
#############
###Model VECM
#############
Full sample size: 2344       End sample size: 2336
Number of variables: 2       Number of estimated slope parameters 32
AIC 30324.45        BIC 30514.4        SSR 14075304
Cointegrating vector (estimated by ML):
   ETH        BNB
r1   1 -6.312237


             ECT                 Intercept          ETH -1             BNB -1
Equation ETH -0.0064(0.0043)     2.2890(1.7623)     0.0015(0.0292)     -0.5466(0.1876)**
Equation BNB 0.0013(0.0007).     0.0269(0.2736)     0.0048(0.0045)     -0.1315(0.0291)***
             ETH -2              BNB -2             ETH -3              BNB -3
Equation ETH -0.0191(0.0289)     0.2277(0.1875)     0.0226(0.0289)      0.2783(0.1873)
Equation BNB -0.0053(0.0045)     0.1068(0.0291)***  -0.0004(0.0045)     0.0569(0.0291).
             ETH -4             BNB -4              ETH -5
Equation ETH 0.0387(0.0289)     -0.0804(0.1862)     -0.0300(0.0290)
Equation BNB 0.0212(0.0045)***  -0.0998(0.0289)***  -0.0143(0.0045)**
             BNB -5              ETH -6             BNB -6             ETH -7
Equation ETH -0.2241(0.1869)     0.1565(0.0291)*** -0.7702(0.1862)*** -0.0231(0.0293)
Equation BNB -0.0252(0.0290)     0.0100(0.0045)*   0.0132(0.0289)     -0.0025(0.0045)
             BNB -7
Equation ETH 0.1478(0.1868)
Equation BNB 0.0364(0.0290)
```

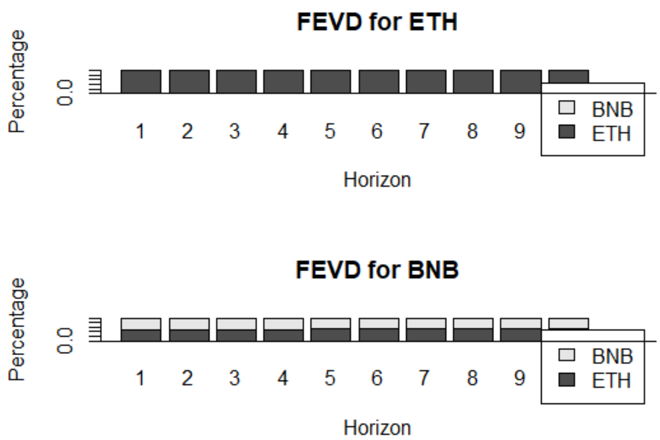
Each lagged variable's effect on the current value of its corresponding variable is represented by the coefficients. The cointegrating vector shows how the variables have interconnected Ethereum and BNB coins throughout time.

Then we plotted the impulse response function which is a graphical representation of the response of a variable to a one-time shock to another variable in a dynamic system.



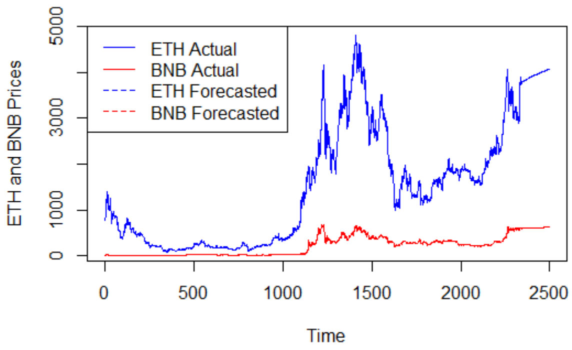
The Ethereum line varies just above suggesting that the initial shock to BNB's price was followed by a somewhat beneficial reaction over time. This IRF helps analyze how BNB price shocks impact Ethereum's price. The IRF plot provides useful information about the dynamic relationships between the variables in a system and can be used to answer questions like, how does a shock to one variable affect the other variables in the system? What is the speed of adjustment of the variables in response to the shock? and how persistent is the response of each variable to the shock?

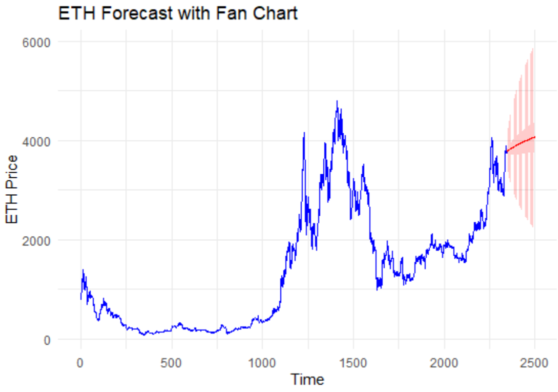
Then we plotted variance decomposition it shows proportion of movement in the Ethereum price



The graph shows the results of a Forecast Error Variance Decomposition. It shows how much of the unpredictable error in forecasting one variable is due to variations in its past values and the past values of the other variable in the model.

Model forecasting





The first graph combines predicted and actual figures to show how the prices of BNB and Ethereum are expected to behave. The anticipated values and prediction intervals are shown in a fan chart, which effectively visualizes the forecast uncertainty. Ethereum's price-rising trend is represented by a wide red region on the graph in the forecasting period, emphasizing the increased level of uncertainty around it.

```
ETH Accuracy Metrics:
MAPE: 12.47762 %
```

Since the Ethereum price Mean Absolute Percentage Error is less than 15% shows high accuracy.

References

Binance. (2024, 06 01). Retrieved 06 01, 2024, from https://www.binance.com/en/blog/all/binance-research-bnb-performed-best-btc-dominated-market-in-2019-421499824684900375