FINANCIAL ANALYTICS

ASSIGNMENT

Name: Khushi Malik
Roll no. 15A

1. Introduction

This report presents the forecasting and analysis of cryptocurrency prices for **DOGE-USD** and **Ethereum USD** using RStudio. The primary goal is to explore the historical trends of these cryptocurrencies, perform a time series analysis, and predict future values using statistical techniques. The analysis uses **descriptive statistics (time series plots)** and the **ARIMA** model for forecasting.

2. Data Overview

The dataset for this analysis spans from December 31st, 2019, to October 7th, 2024. The price data for both DOGE-USD and Ethereum USD was sourced from Investing.com, a trusted platform for financial data. Over this period, the dataset captures daily price fluctuations for both cryptocurrencies, allowing us to analyze their performance over almost five years.

The time series data reflects key trends, volatility, and market responses to various global economic and industry-specific events. Using this comprehensive range, we aim to build accurate forecasting models that consider historical movements and potential future trends.

3. Descriptive Statistics

Descriptive statistics provide a summary of data characteristics, helping to quantify and present important features like central tendency (mean, median) and variability (standard deviation) of datasets.

1. **BasicStats**: It is used to compute descriptive statistics such as mean, median, standard deviation, and quartiles, which summarize the central tendency, variability, and distribution of the data.
2. **Tsplot**: In time series analysis, **tsplot** provides visualization of the data, offering insights into trends, seasonality, and other time-dependent patterns. These plots are often accompanied by basic statistics to analyse patterns over time.

BasicStats helps provide statistical measures of past prices, while **tsplot** helps visualize the price changes and trends over time.

3.1 DOGE-USD

```
> basicStats(Dogecoin)
                Dogecoin
nobs        1743.000000
NAs            0.000000
Minimum        0.001585
Maximum        0.686880
1. Quartile    0.055887
3. Quartile    0.138052
Mean           0.099645
Median         0.078640
Sum          173.680413
SE Mean        0.002178
LCL Mean       0.095372
UCL Mean       0.103917
Variance       0.008271
Stdev          0.090946
Skewness       1.743542
Kurtosis       5.033092
```

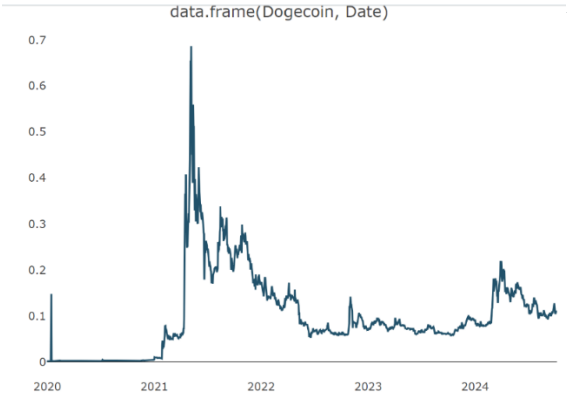
```
> basicStats(LRDogecoin)
                LRDogecoin
nobs        1742.000000
NAs            0.000000
Minimum       -4.141786
Maximum        4.155892
1. Quartile   -0.020367
3. Quartile    0.024227
Mean          -0.002288
Median         0.000030
Sum           -3.985725
SE Mean        0.003817
LCL Mean      -0.009775
UCL Mean       0.005199
Variance       0.025382
Stdev          0.159317
Skewness      -0.516419
Kurtosis     531.246101
```

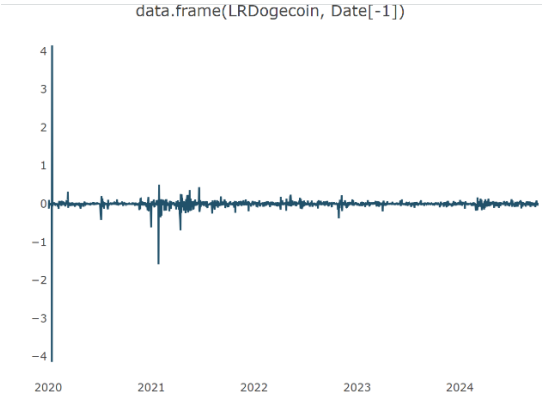
The values range from -4.141786 to 4.155892, suggesting a significant spread in LRDogecoin prices. The **standard deviation** is 0.159317, indicating a relatively high variability in prices.

Shape:

- **Skewness:** The skewness of -0.516419 suggests a slight left skew, meaning there are a few extremely low values that pull the mean to the left.
- **Kurtosis:** The kurtosis of 531.246101 indicates a heavy-tailed distribution, suggesting that extreme values are much more common than in a normal distribution.

Overall, the statistics show that LRDogecoin prices have a wide range, with a significant number of outliers on both the high and low ends. The distribution is slightly skewed to the left, and there is a heavy-tailed distribution, indicating that extreme values are more common than in a normal distribution.





LRDogecoin has experienced significant price volatility throughout the period, indicating a high degree of risk associated with investing in DOGE-USD. There is no clear upward or downward

direction suggesting that the cryptocurrency's value has been influenced by various factors, including market sentiment, technological developments, and economic conditions.

3.2 Ethereum USD

```
> basicStats(Ethereum)
                Ethereum
nobs         1.743000e+03
NAs          0.000000e+00
Minimum      1.079000e+02
Maximum      4.808380e+03
1. Quartile  1.217610e+03
3. Quartile  2.762190e+03
Mean         1.923921e+03
Median       1.840920e+03
Sum          3.353394e+06
SE Mean      2.730856e+01
LCL Mean     1.870360e+03
UCL Mean     1.977482e+03
Variance     1.299855e+06
Stdev        1.140112e+03
Skewness     1.604020e-01
Kurtosis    -7.325530e-01
```

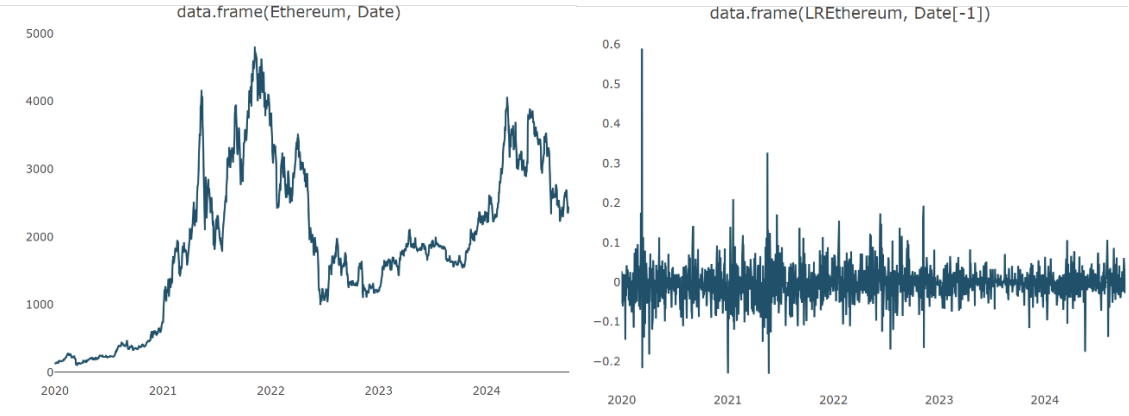
```
> basicStats(LREthereum)
             LREthereum
nobs       1742.000000
NAs           0.000000
Minimum      -0.230772
Maximum       0.589639
1. Quartile  -0.022642
3. Quartile   0.017689
Mean         -0.001683
Median       -0.001585
Sum          -2.931220
SE Mean       0.001085
LCL Mean     -0.003812
UCL Mean      0.000446
Variance      0.002052
Stdev         0.045303
Skewness      1.422085
Kurtosis     20.231232
```

The values range from 0.000000 to 0.589639, suggesting a moderate spread in LREthereum prices. The **standard deviation** is 0.045303, indicating a relatively low variability in prices compared to other cryptocurrencies.

Shape:

- **Skewness:** The skewness of 1.422085 suggests a right-skewed distribution, meaning there are a few extremely high values that pull the mean to the right.
- **Kurtosis:** The kurtosis of 20.231232 indicates a heavy-tailed distribution, suggesting that extreme values are more common than in a normal distribution.

Overall, the statistics show that LREthereum prices have a moderate range, with a significant number of outliers on the higher end. The distribution is skewed to the right, and there is a heavy-tailed distribution, indicating that extreme values are more common than in a normal distribution.



LREthereum USD has experienced significant price volatility throughout the period, indicating a high degree of risk associated with investing in Ethereum USD. There is no clear upward or downward direction suggesting that the cryptocurrency's value has been influenced by various factors, including market sentiment, technological developments, and economic conditions.

4. Econometric Modelling

In the **econometric modelling** of time series data, the **ARIMA (Autoregressive Integrated Moving Average)** model is used to forecast future values based on past observations. ARIMA integrates three components: Autoregression (AR), Differencing (I), and a Moving Average (MA). It can handle non-stationary data by using differencing and is widely used for forecasting cryptocurrency trends, including **DOGECOIN** and **Ethereum USD**.

Model Breakdown:

1. **Autoregressive (AR)**: This refers to the use of past values in the model to predict future values.
2. **Integrated (I)**: This refers to the differencing of raw observations to make the time series stationary.
3. **Moving Average (MA)**: This involves modeling the error term as a linear combination of error terms occurring at previous times.

General ARIMA Equation:

For an **ARIMA(p,d,q)** model, the equation is:

$$y'_t = I + \alpha_1 y'_{t-1} + \alpha_2 y'_{t-2} + \cdots + \alpha_p y'_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$

Where:

- Yt is the forecasted value at time **t**.
- **p** is the number of lag terms (AR part).
- **d** is the degree of differencing (I part).
- **q** is the number of lagged forecast errors (MA part).
- α are the coefficients for the AR part.
- θ are the coefficients for the MA part.
- et is the white noise error term.

For the cryptocurrency data analysis (DOGECOIN and Ethereum USD USD), the specific ARIMA model would be determined based on the best fit for the time series after conducting tests for stationarity and model selection using criteria like **AIC** (Akaike Information Criterion).

Forecasting Application:

In this report, ARIMA models were applied using **RStudio** to forecast DOGECOIN and Ethereum USD prices. The model parameters (**p**, **d**, and **q**) were determined based on the autocorrelation and partial autocorrelation functions (ACF and PACF) from the time series data.

4.1 DOGE-USD

```
> adf.test(LRDogecoin)

        Augmented Dickey-Fuller Test

data:  LRDogecoin
Dickey-Fuller = -12.208, Lag order = 12, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(LRDogecoin) : p-value smaller than printed p-value
```

Interpretation:

- **p-value:** The p-value of 0.01 is less than the commonly used significance level of 0.05. This indicates that we can reject the null hypothesis of non-stationarity in favour of the alternative hypothesis of stationarity.
- **Stationarity:** Therefore, based on the ADF test results, we can conclude that the LRDogecoin time series data is likely stationary. This means that its statistical properties, such as mean and variance, are relatively constant over time.

```
> auto.arima(LRDogecoin)
Series: LRDogecoin
ARIMA(0,0,1) with zero mean

Coefficients:
          ma1
      -0.4783
s.e.   0.0205

sigma^2 = 0.02056:  log likelihood = 911.77
AIC=-1819.55   AICc=-1819.54   BIC=-1808.62
```

Interpretation:

- **ARIMA(0,0,1):** This indicates that the model is a simple moving average (MA) model with no autoregressive (AR) terms and no differencing (I) terms. The MA term with a lag of 1 suggests that the current value of LRDogecoin is influenced by the previous value.
- **Coefficients:** The coefficient for the MA term (ma1) is -0.4783, which means that the current value of LRDogecoin is negatively correlated with the previous value. A negative coefficient indicates that if the previous value was high, the current value is likely to be lower.
- **Standard error:** The standard error for the MA coefficient is 0.0205, which is relatively small. This suggests that the coefficient is statistically significant and provides a meaningful contribution to the model.
- **Sigma^2:** This is the estimated variance of the error term in the model. A lower value indicates a better fit of the model to the data.
- **Log likelihood, AIC, AICc, BIC:** These are information criteria used to compare the fit of different models. Lower values generally indicate a better-fitting model.

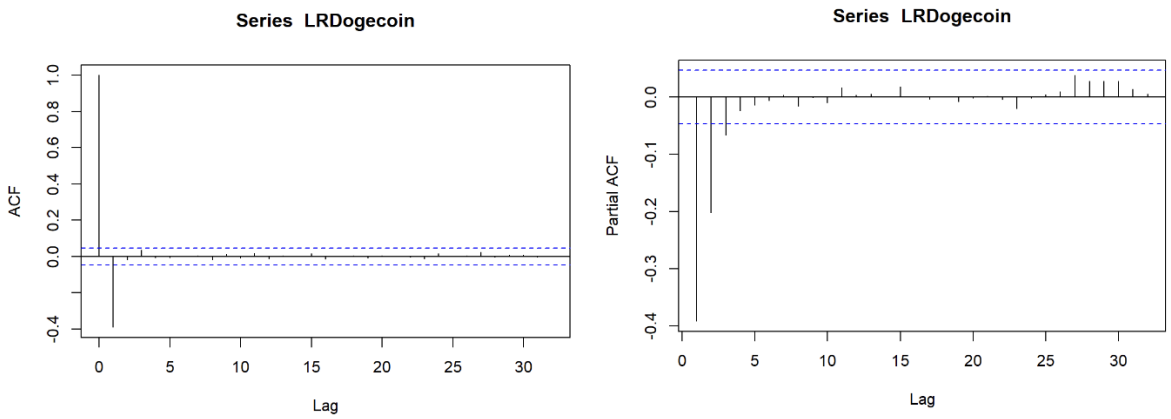
Equation:

The ARIMA(0,0,1) model can be expressed as:

LRDOGECOIN(t) = -0.4783 * LRDOGECOIN(t-1) + ε(t)

where:

- LRDOGECOIN(t) is the current value of LRDOGECOIN.
- LRDOGECOIN(t-1) is the previous value of LRDOGECOIN.
- ε(t) is the error term, which is assumed to be white noise.



The **ACF plot** shows a significant negative autocorrelation at lag 1, indicating a mean-reverting behaviour in LRDOGECOIN. There is no significant autocorrelation at higher lags, suggesting that the MA(1) term in the ARIMA model is adequate.

The **PACF plot** shows a significant negative partial autocorrelation at lag 1, confirming the mean-reverting behaviour in LRDOGECOIN. There is no significant partial autocorrelation at higher lags, suggesting that the MA(1) term in the ARIMA model is adequate.

```
> auto.arima(LRDogecoin)
Series: LRDogecoin
ARIMA(0,0,1) with zero mean

Coefficients:
          ma1
      -0.4783
s.e.   0.0205

sigma^2 = 0.02056:  log likelihood = 911.77
AIC=-1819.55   AICc=-1819.54   BIC=-1808.62
```

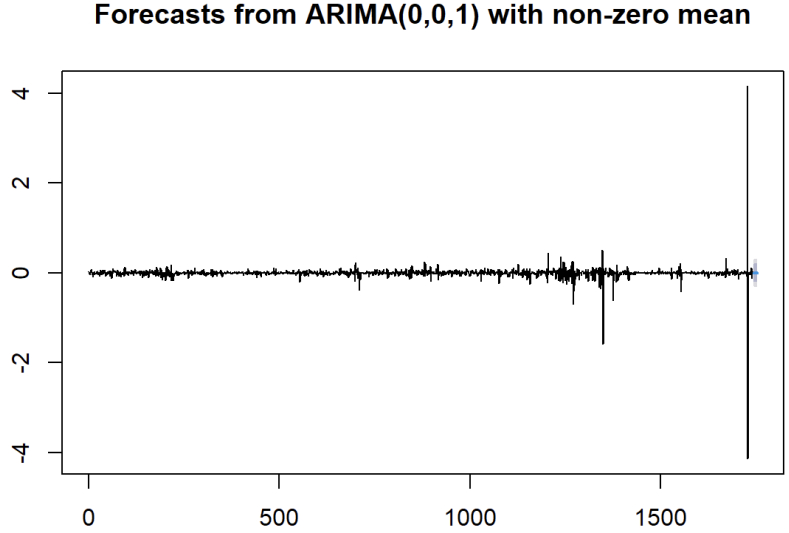
```
> forecast(model1,10)
     Point Forecast      Lo 80     Hi 80      Lo 95     Hi 95
1743    0.003516364 -0.1801162 0.1871489 -0.2773254 0.2843581
1744   -0.002289209 -0.2059494 0.2013710 -0.3137606 0.3091821
1745   -0.002289209 -0.2059494 0.2013710 -0.3137606 0.3091821
1746   -0.002289209 -0.2059494 0.2013710 -0.3137606 0.3091821
1747   -0.002289209 -0.2059494 0.2013710 -0.3137606 0.3091821
1748   -0.002289209 -0.2059494 0.2013710 -0.3137606 0.3091821
1749   -0.002289209 -0.2059494 0.2013710 -0.3137606 0.3091821
1750   -0.002289209 -0.2059494 0.2013710 -0.3137606 0.3091821
1751   -0.002289209 -0.2059494 0.2013710 -0.3137606 0.3091821
1752   -0.002289209 -0.2059494 0.2013710 -0.3137606 0.3091821
```

Columns:

- **Point Forecast:** The predicted value for the corresponding period.
- **Lo 80:** The lower bound of the 80% confidence interval for the forecast.
- **Hi 80:** The upper bound of the 80% confidence interval for the forecast.
- **Lo 95:** The lower bound of the 95% confidence interval for the forecast.
- **Hi 95:** The upper bound of the 95% confidence interval for the forecast.

Interpretation:

- **Point Forecast:** The forecasted values for LRDOGECOIN range from -0.002289209 to 0.003516364.
- **Confidence Intervals:** The 80% and 95% confidence intervals provide a range of plausible values for the forecasted LRDOGECOIN prices. A wider confidence interval indicates greater uncertainty in the forecast.
- **Overall:** The model predicts that LRDOGECOIN will remain relatively stable in the short term, with a slight negative bias. However, there is a degree of uncertainty associated with the forecasts, as indicated by the confidence intervals. It's important to note that these are just predictions, and the actual values may deviate from the forecasts.



4.2 ETHEREUM USD

```
> adf.test(LREthereum)

        Augmented Dickey-Fuller Test

data:  LREthereum
Dickey-Fuller = -11.051, Lag order = 12, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(LREthereum) : p-value smaller than printed p-value
```

Interpretation:

- **p-value:** The p-value of 0.01 is less than the commonly used significance level of 0.05. This indicates that we can reject the null hypothesis of non-stationarity in favour of the alternative hypothesis of stationarity.
- **Stationarity**: Therefore, based on the ADF test results, we can conclude that the LREthereum USD time series data is likely stationary. This means that its statistical properties, such as mean and variance, are relatively constant over time.

```
> auto.arima(LREthereum)
Series: LREthereum
ARIMA(1,0,1) with non-zero mean

Coefficients:
          ar1     ma1     mean
      -0.7729  0.7008  -0.0017
s.e.   0.0839  0.0940   0.0010

sigma^2 = 0.002029:  log likelihood = 2930.3
AIC=-5852.59   AICc=-5852.57   BIC=-5830.74
```

Interpretation:

- **ARIMA(1,0,1):** This indicates that the model is an ARMA(1,1) model with a non-zero mean. The AR term with a lag of 1 suggests that the current value of LREthereum USD is influenced by the previous value, and the MA term with a lag of 1 suggests that the current value is also influenced by the previous error term.
- **Coefficients:**
    - The coefficient for the AR term (ar1) is -0.7729, which means that the current value of LREthereum USD is negatively correlated with the previous value. A negative coefficient indicates that if the previous value was high, the current value is likely to be lower.
    - The coefficient for the MA term (ma1) is 0.7008, which means that the current value of LREthereum USD is positively correlated with the previous error term. A positive coefficient indicates that if the previous error was positive (the previous value was overpredicted), the current value is likely to be higher.
    - The mean term is -0.0017, which represents the constant term in the model. It suggests that the average value of LREthereum USD is slightly negative.

- **Standard errors:** The standard errors for the coefficients are relatively small, suggesting that the coefficients are statistically significant and provide meaningful contributions to the model.
- **Sigma^2:** This is the estimated variance of the error term in the model. A lower value indicates a better fit of the model to the data.
- **Log likelihood, AIC, AICc, BIC:** These are information criteria used to compare the fit of different models. Lower values generally indicate a better-fitting model.

Equation:

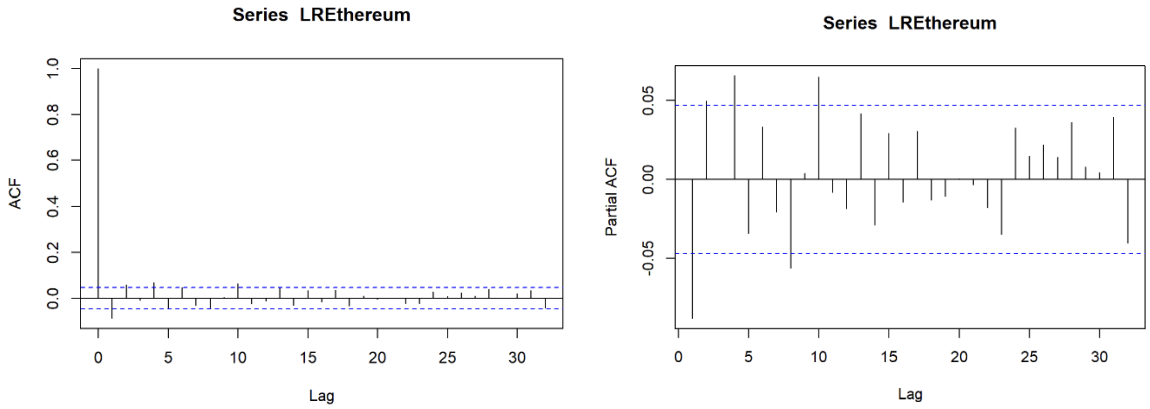
The ARIMA(1,0,1) model can be expressed as:

LREthereum USD USD(t) = -0.7729 * LREthereum USD USD(t-1) + 0.7008 * ε(t-1) - 0.0017 + ε(t)

where:

- LREthereum USD USD(t) is the current value of LREthereum USD.
- LREthereum USD USD(t-1) is the previous value of LREthereum USD.
- ε(t) is the error term, which is assumed to be white noise.
- ε(t-1) is the previous error term.

This equation shows that the current value of LREthereum USD is predicted based on a weighted average of the previous value and the previous error term, with the weights given by the coefficients for the AR and MA terms. Additionally, the constant term of -0.0017 represents the expected average value of LREthereum USD.



The **ACF plot** shows a significant positive autocorrelation at lag 1, suggesting a momentum effect in LREthereum USD. There is no significant autocorrelation at higher lags, indicating that the ARMA(1,1) terms in the ARIMA model are adequate.

The **PACF plot** shows a significant positive partial autocorrelation at lag 1, confirming the momentum effect in LREthereum USD. There is no significant partial autocorrelation at higher lags, indicating that the ARMA(1,1) terms in the ARIMA model are adequate.

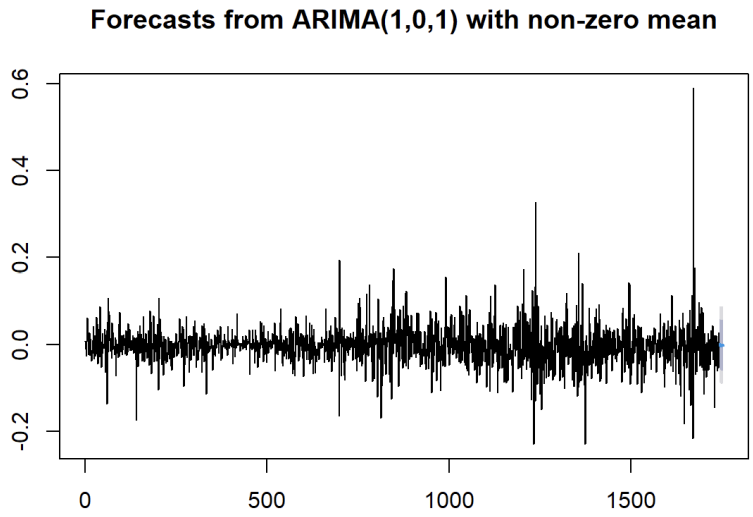
```
> model2 = arima(LREthereum , order = c(1,0,1))
> summary(model2)

Call:
arima(x = LREthereum, order = c(1, 0, 1))

Coefficients:
          ar1     ma1  intercept
      -0.7729  0.7008    -0.0017
s.e.   0.0839  0.0940     0.0010

sigma^2 estimated as 0.002025:  log likelihood = 2930.3,  aic = -5852.59

Training set error measures:
                      ME       RMSE        MAE      MPE     MAPE
Training set 2.439987e-05 0.04500009 0.03016512 96.57773 124.9843
                   MASE         ACF1
Training set 0.6587367 -0.00811562

> forecast(model2,10)
     Point Forecast       Lo 80      Hi 80       Lo 95      Hi 95
1743   0.0014156874 -0.05625424 0.05908562 -0.08678286 0.08961423
1744  -0.0041171692 -0.06193680 0.05370246 -0.09254466 0.08431032
1745   0.0001592768 -0.05774960 0.05806815 -0.08840471 0.08872326
1746  -0.0031460665 -0.06110819 0.05481606 -0.09179149 0.08549936
1747  -0.0005913062 -0.05858522 0.05740261 -0.08928535 0.08810274
1748  -0.0025659272 -0.06057883 0.05544697 -0.09128900 0.08615715
1749  -0.0010397065 -0.05906394 0.05698453 -0.08978012 0.08770071
1750  -0.0022193504 -0.06025036 0.05581166 -0.09097012 0.08653142
1751  -0.0013075821 -0.05934264 0.05672747 -0.09006454 0.08744937
1752  -0.0020123045 -0.06004977 0.05602517 -0.09077296 0.08674835
```

Columns:

- **Point Forecast**: The predicted value for the corresponding period.
- **Lo 80:** The lower bound of the 80% confidence interval for the forecast.
- **Hi 80:** The upper bound of the 80% confidence interval for the forecast.
- **Lo 95:** The lower bound of the 95% confidence interval for the forecast.
- **Hi 95:** The upper bound of the 95% confidence interval for the forecast.

Interpretation:

- **Point Forecast**: The forecasted values for LREthereum USD range from -0.0041171692 to 0.0031460665.
- **Confidence Intervals:** The 80% and 95% confidence intervals provide a range of plausible values for the forecasted LREthereum USD prices. A wider confidence interval indicates greater uncertainty in the forecast.
- **Overall:** The model predicts that LREthereum USD will remain relatively stable in the short term, with a slight downward bias. However, there is a degree of uncertainty associated with the forecasts, as indicated by the confidence intervals. It's important to note that these are just predictions, and the actual values may deviate from the forecasts.



5. Results:

1. **Descriptive Statistics:**
    a. **DOGE-USD**: Prices range from -4.141786 to 4.155892, with a high standard deviation of 0.159317, indicating high volatility. The distribution is slightly skewed to the left with heavy tails, showing frequent extreme values.
    b. **Ethereum USD:** Prices range from 0 to 0.589639 with lower volatility (standard deviation of 0.045303). The distribution is skewed to the right, and it also exhibits heavy tails with frequent extreme values.
2. **Econometric Modelling (ARIMA):**
    a. **DOGE-USD (ARIMA 0,0,1):** A simple moving average model with significant autocorrelation at lag 1. The model predicts mean-reverting behaviour with a slight negative bias.
    b. **Ethereum USD (ARIMA 1,0,1):** An ARMA model that shows positive momentum with both autoregressive and moving average terms. The model suggests Ethereum USD prices are influenced by the previous value and error.

6. Conclusion:

Both cryptocurrencies demonstrate high volatility with no clear long-term trend. ARIMA models provided reasonable forecasts but with uncertainty, as indicated by confidence intervals. The DOGE-USD model predicts slight downward stability, while the Ethereum USD model forecasts slight negative momentum in the short term.

7. Reference

Yadav, M. P., Sehgal, V., Ratra, D., & Wajid, A. (2023). Forecasting the energy commodities: an evidence of ARIMA and intervention analysis. *International Journal of Monetary Economics and Finance*, *16*(6), 443–457. https://doi.org/10.1504/ijmef.2023.136086