LINK-USDT Thinking logic

1. Why Conflicting Results from ADF and KPSS?

ADF tests for a unit root (non-stationarity) as the null hypothesis.

KPSS tests for stationarity as the null hypothesis.

These two tests can give conflicting results because they are based on different statistical assumptions. This is why it's useful to combine them with visual analysis or other tools.

2. Output:

```
> D_seas
> d_adf
[1] 1
> d_kpss
[1] 0
> D_seas
[1] 0
```

d_adf = 1 vs. d_kpss = 0:

- The ADF test suggests the series is non-stationary and requires one difference to become stationary.
- The KPSS test suggests the series is already stationary.
- These conflicting results are common and highlight the importance of examining the data visually (e.g., via plots) or using additional tests.

D_seas = 0:

- Seasonal differencing is not needed, so there is no evidence of strong seasonality in the data.

3. RESULT:

```
        Augmented Dickey-Fuller Test

data:  na.omit(link_diff)
Dickey-Fuller = -3.1346, Lag order = 3, p-value = 0.1162
alternative hypothesis: stationary
```

The p-value (0.1162) is greater than the standard significance level (e.g., 0.05).

Conclusion: Fail to reject the null hypothesis that the differenced data link_diff is non-stationary.

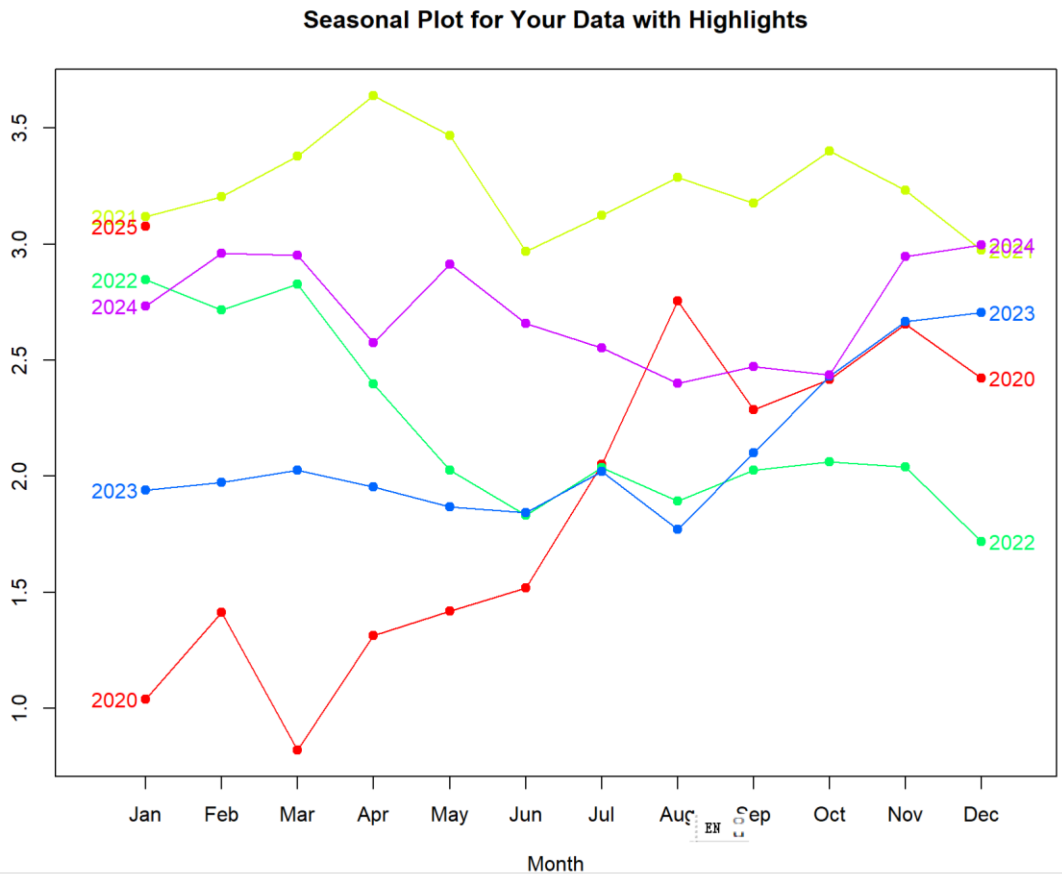
```
> kpss_1

        KPSS Test for Level Stationarity

data:  na.omit(link_diff)
KPSS Level = 0.18043, Truncation lag parameter = 3, p-value = 0.1
```

The p-value (0.1) is above the 0.05 significance threshold.

Conclusion: Fail to reject the null hypothesis that the data is stationary.

While the KPSS Shows stationary but ADF not shows.

4. RESULT: seasonal plot with log price

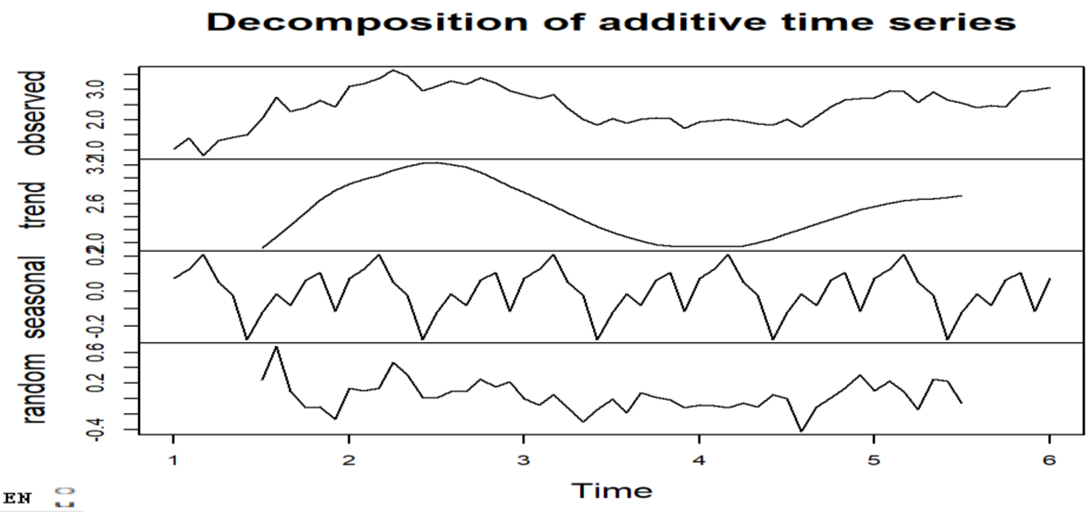


Seasonality Is Weak but Present:

- Both plots indicate **some level of seasonality**, with recurring higher values in the **summer months (June to August)** and lower values in the **winter months (January and February)**.
- However, the seasonal effect is **not consistent across all years**, as seen in 2024, where the trend declines mid-year instead of peaking.

5. Result

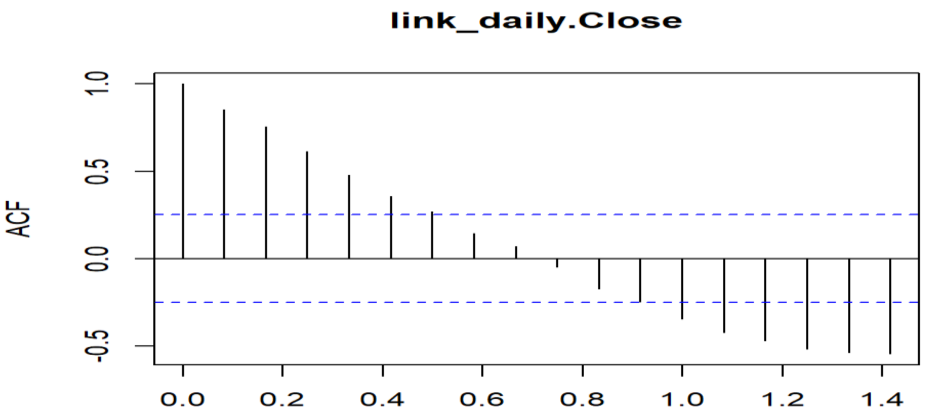
Perform decomposition to separate the trend, seasonality, and residuals for clearer insights.



Trend: The data is primarily driven by a long-term upward trend.

Seasonality: There is a weak seasonal component.

Randomness: The residuals suggest no significant structure beyond the trend and seasonality.



The data is **non-stationary** due to the strong trend component.

Seasonality is weak or absent, as shown by the lack of significant spikes at seasonal lags in the ACF.

To analyze and model this data, apply **differencing** to remove the trend and consider fitting a SARIMA model if weak seasonality is confirmed.

6. Result : try auto-arima to find seasonality and good arima model but really bad result

```
> summary(auto_sarima)
Series: link_log
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1    mean
      0.9247  2.3260
s.e.  0.0514  0.3989

sigma^2 = 0.07814:  log likelihood = -8.75
AIC=23.51   AICc=23.93   BIC=29.84

Training set error measures:
                     ME      RMSE       MAE        MPE   MAPE      MASE       ACF1
Training set 0.03258469 0.2749233 0.2167849 -0.2910554 10.376 0.2509808 -0.1430269
```

AR(1) Coefficient = 0.9247:

Indicates strong short-term dependence (the current value is highly influenced by the previous value).

Model Selection Criteria

AIC (23.51), AICc (23.93), BIC (29.84):

These values indicate how well the model fits the data while penalizing for complexity.

Training Set Error Measures

Mean Absolute Percentage Error (MAPE) = 10.376:

A MAPE of ~10% indicates the model predictions are off by an average of 10%. While acceptable for some cases, lower values (e.g., <5%) are preferred for better accuracy.

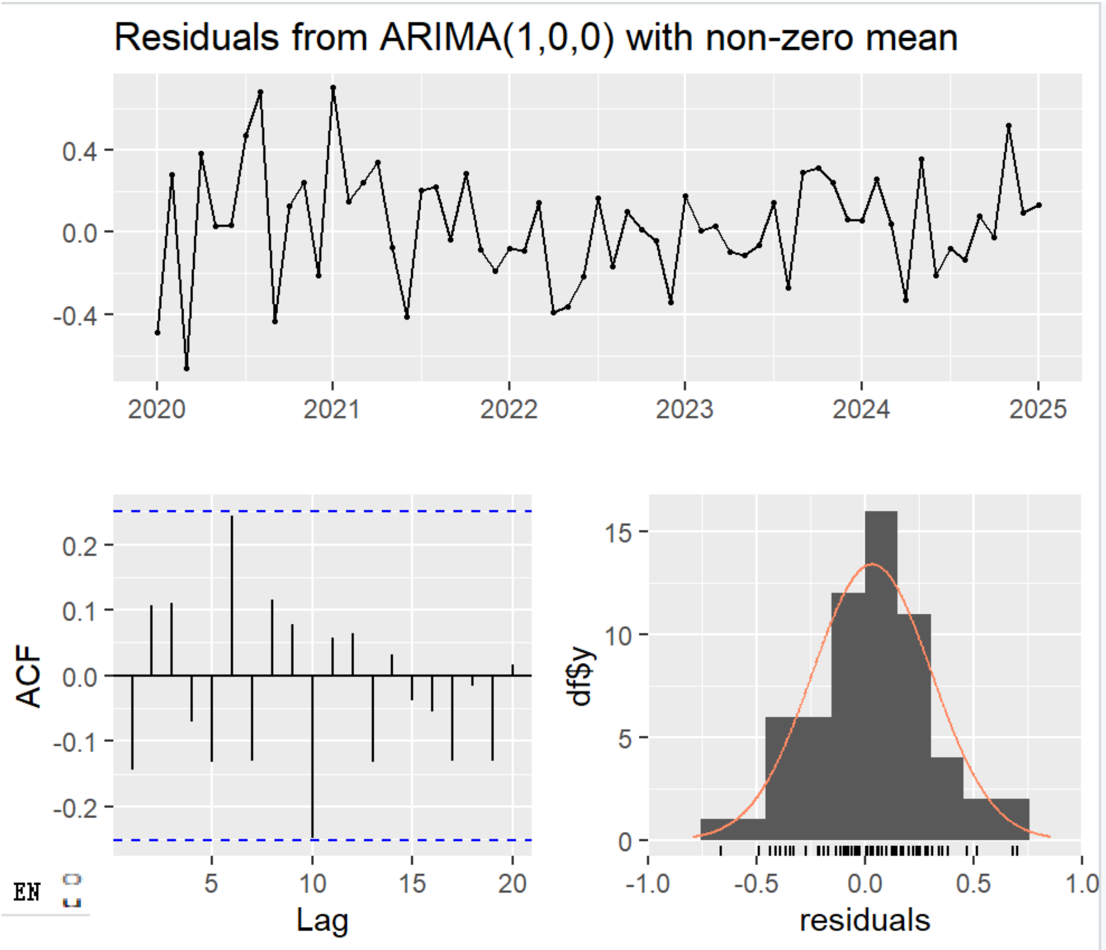
Root Mean Square Error (RMSE) = 0.2749:

Reflects the magnitude of prediction errors. Smaller values indicate better model fit.

ACF1 = -0.1430:

The residuals (errors) from the model show low autocorrelation at lag 1, meaning the model has successfully removed much of the structure in the data.

7. RESULT :I checking the residual of ARIMA(1,0,0)



Strengths:

1.The residuals fluctuate around zero, suggesting the model captures the central tendency well.

2.The residuals' distribution is approximately normal, which is desirable for predictive accuracy.

Weaknesses:

1.The spike in the ACF at lag 5 indicates some **remaining structure** in the residuals that the model failed to capture.

2.Weak periodicity in the time plot of residuals suggests that some seasonal or cyclic components might still exist.

Conclusion

The ARIMA(1,0,0) model works reasonably well but does not fully capture the structure in the data (e.g., lag 5 autocorrelation).

8 . Result In here , we try another way , I based on choosing the best AIC to find model to see if it is possible (I select 48 model in here), and I set an another cycle code (for p and q from 0 to 3) in autoregressive term, and (for p and q for 0 to 2) to find the seasonal autoregressive term. Here are the result:

```
> summary(final_sarima)
Series: link_log
ARIMA(0,1,0)(0,0,1)[12]

Coefficients:
        sma1
      0.2858
s.e.  0.1677

sigma^2 = 0.07447:  log likelihood = -7.22
AIC=18.44   AICc=18.65   BIC=22.63

Training set error measures:
                   ME      RMSE       MAE       MPE     MAPE      MASE       ACF1
Training set 0.025503 0.2683746 0.2038558 0.6382588 9.450273 0.2360123 -0.1104558
```

```
> checkresiduals(final_sarima)

        Ljung-Box test

data:  Residuals from ARIMA(0,1,0)(0,0,1)[12]
Q* = 14.997, df = 11, p-value = 0.1826

Model df: 1.   Total lags used: 12
```

Strengths:

Seasonality Captured:

The inclusion of the seasonal moving average term (SMA1) has improved the model, reducing the AIC and addressing weak seasonality.

Well-Behaved Residuals:

The residuals are largely uncorrelated (confirmed by ACF and Ljung-Box test) and approximately normally distributed.

Error Metrics Are Acceptable:

A MAPE of 9.45% is reasonable for many forecasting applications, though it could be further reduced.

Weaknesses:

Remaining Structure at Lag 5:

The spike in the residual ACF at lag 5 suggests there may be some minor structure not captured by the current model.

All in all what we can find ARIMA(0,1,0)(0,0,1)[12] is still the best by comparing many times with lower aic and bic, also the lower p-value, but maybe because of high volatility we still cannot see this is the best model, and we are going to find a good garch model and arch effect in the next step.

9.Result here:We find no arch effect here, which means maybe seasonal is not the best.

10 . Ressult here :Change the mind to find daily data and arma-garch, result with daily data and best arima selection testing by autoarima with seasonal = false, the answer is arima(1,0,1)

```
Series: link_ret
ARIMA(1,0,1) with zero mean

Coefficients:
          ar1     ma1
      -0.7155  0.6589
s.e.   0.1219  0.1307

sigma^2 = 0.0033:  log likelihood = 2628.13
AIC=-5250.27   AICc=-5250.25   BIC=-5233.74
```

```
> Box.test(residuals(arma_search))

        Box-Pierce test

data:  residuals(arma_search)
X-squared = 0.10483, df = 1, p-value = 0.7461
```

EXPLAIN : **Box-Pierce (or Ljung-Box) Test** on the residuals shows **p = 0.7461** (not significant), meaning **no leftover linear autocorrelation** after fitting ARIMA(1,0,1). So the *mean* dynamics are well captured.

```
> ArchTest(residuals(arma_search))

        ARCH LM-test; Null hypothesis: no ARCH effects

data:  residuals(arma_search)
Chi-squared = 63.412, df = 12, p-value = 5.352e-09
```

EXPLAIN: **ARCH LM-Test** on the same residuals has **p = 5.35e-09** (very significant), strongly indicating **volatility clustering** or time-varying variance.

11. RESULT:

here I write a loop for garch(p,q) ,p from [1:3]and q from[1:3] to find the best garch model by comparing the "sGARCH", "eGARCH", "gjrGARCH", in Student-t, Normal, Skewed Student-t distribution. Which means we have already detected 81 ways to find the best model, we can still do more literation. But now based on 81 data comparing

in the end we find it is basic symmetric sgarch (1,1) model and it residual follow student t-distributon

```
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model      : sGARCH(1,1)
Mean Model       : ARFIMA(1,0,1)
Distribution     : std

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.001479    0.000905   1.6347 0.102108
ar1     0.723020    0.276713   2.6129 0.008978
ma1    -0.753913    0.262416  -2.8730 0.004066
omega   0.000115    0.000040   2.8782 0.003999
alpha1  0.102946    0.022575   4.5602 0.000005
beta1   0.864631    0.028930  29.8873 0.000000
shape   5.219507    0.613517   8.5075 0.000000

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.001479    0.000914   1.6179 0.105688
ar1     0.723020    0.289072   2.5012 0.012378
ma1    -0.753913    0.272932  -2.7623 0.005740
omega   0.000115    0.000051   2.2517 0.024343
alpha1  0.102946    0.027807   3.7022 0.000214
beta1   0.864631    0.038279  22.5875 0.000000
shape   5.219507    0.695308   7.5068 0.000000

LogLikelihood : 2876.413
```

Analysis of Current Model

Model Fit (Parameters & Information Criteria):

ARMA(1,1) Mean Model:

AR(1) and MA(1) are statistically significant with low ppp-values, meaning the mean model is capturing some patterns in the data.

sGARCH(1,1):

Parameters (α1,β1) are statistically significant, with β1=0.864 suggesting strong volatility persistence.

ω (constant term) is significant but very small, indicating low baseline volatility.

12. Result in AIC/BIC

```
Information Criteria
------------------------------------

Akaike       -3.1411
Bayes        -3.1200
Shibata      -3.1411
Hannan-Quinn -3.1333
```

Information Criteria:

AIC: −3.1411-3.1411−3.1411, BIC: −3.1200-3.1200−3.1200, and Log-Likelihood: 2876.412876.412876.41. These are reasonable for this type of model, but improvement is always worth exploring.

13. RESULT BY RESIDUAL Diagnostics:

```
Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic   p-value
Lag[1]                      0.397 5.287e-01
Lag[2*(p+q)+(p+q)-1][5]     6.187 3.616e-05
Lag[4*(p+q)+(p+q)-1][9]     9.089 2.200e-02
d.o.f=2
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.0143  0.9048
Lag[2*(p+q)+(p+q)-1][5]    0.4631  0.9632
Lag[4*(p+q)+(p+q)-1][9]    1.1771  0.9777
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]    0.4625 0.500 2.000  0.4965
ARCH Lag[5]    0.7137 1.440 1.667  0.8193
ARCH Lag[7]    1.0950 2.315 1.543  0.8976
```

Ljung-Box Test on Residuals:

For raw residuals, $p>0.05$ indicates no significant autocorrelation.

For squared residuals, $p>0.05$ also suggests no leftover volatility clustering.

These results confirm the model has adequately captured the dynamics of the data.

ARCH-LM Test:

$p>0.05$ at all lags indicates no remaining ARCH effects in the residuals.

This means the GARCH component effectively captured volatility clustering.

```
Sign Bias Test
------------------------------------
                   t-value   prob sig
Sign Bias          0.28477 0.7759    
Negative Sign Bias 0.64071 0.5218    
Positive Sign Bias 0.01973 0.9843    
Joint Effect       1.12677 0.7706    


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     16.97       0.5916
2    30     25.23       0.6661
3    40     35.31       0.6389
4    50     42.05       0.7487
```

Sign Bias & Goodness-of-Fit Tests:

Sign Bias Test:

Insignificant p-values suggest no asymmetry in the model residuals, which aligns with using a symmetric distribution (std).

Goodness-of-Fit Test:

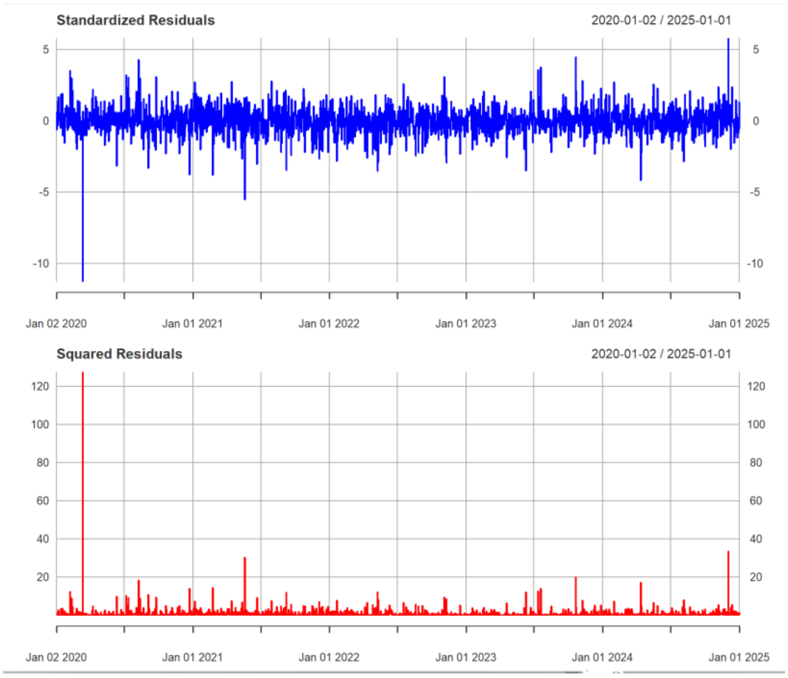
High p-values indicate the model fits the data well across multiple groups.

```
Nyblom stability test
------------------------------------
Joint Statistic:  1.9612
Individual Statistics:             
mu     0.54686
ar1    0.05288
ma1    0.04751
omega  0.77127
alpha1 0.74309
beta1  0.82149
shape  0.33770

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:         1.69 1.9 2.35
Individual Statistic:    0.35 0.47 0.75
```

Nyblom Stability Test:

The joint statistic is below critical values, indicating parameter stability over time. No evidence of structural breaks.

14: Stationarity Analysis



Standardized Residuals:

The residuals appear to fluctuate around zero without any visible trend, which is a key indicator of stationarity.

There are some extreme outliers, but overall, the fluctuations seem consistent over time.

the mean and variance do not appear to change significantly, which supports the assumption of stationarity.

Squared Residuals:

These reflect the volatility clustering (variance changes over time).

Although the squared residuals generally exhibit small spikes, there are periods of increased volatility, especially at the beginning and end.

This could suggest that the volatility is modeled well, but the extreme spikes might indicate some unmodeled dynamics.

```
> par(mfrow = c(1, 1))
> adf.test(link_ret)

        Augmented Dickey-Fuller Test

data:  link_ret
Dickey-Fuller = -12.057, Lag order = 12, p-value = 0.01
alternative hypothesis: stationary
```

```
> kpss.test(link_ret)

        KPSS Test for Level Stationarity

data:  link_ret
KPSS Level = 0.20485, Truncation lag parameter = 8, p-value = 0.1
```

KPSS Statistic:

0.204850. which is low and well within the critical values for the null hypothesis.

p-value:

0.10., indicating that we fail to reject the null hypothesis of stationarity at the 10% significance level.

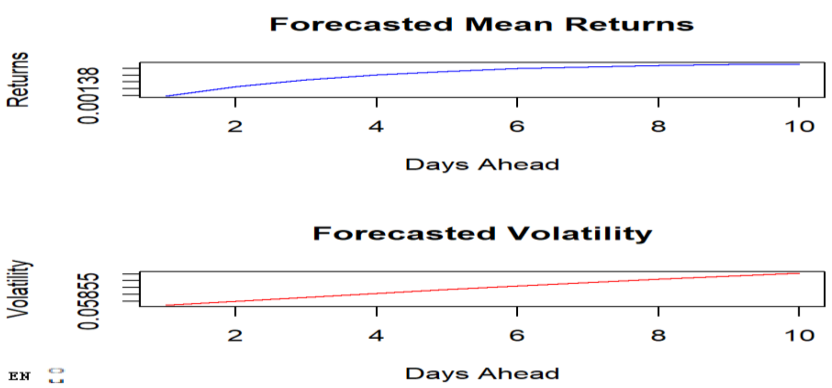
Null Hypothesis:

The null hypothesis of the KPSS test is that the series is stationary around a constant level or deterministic trend

15. prediction

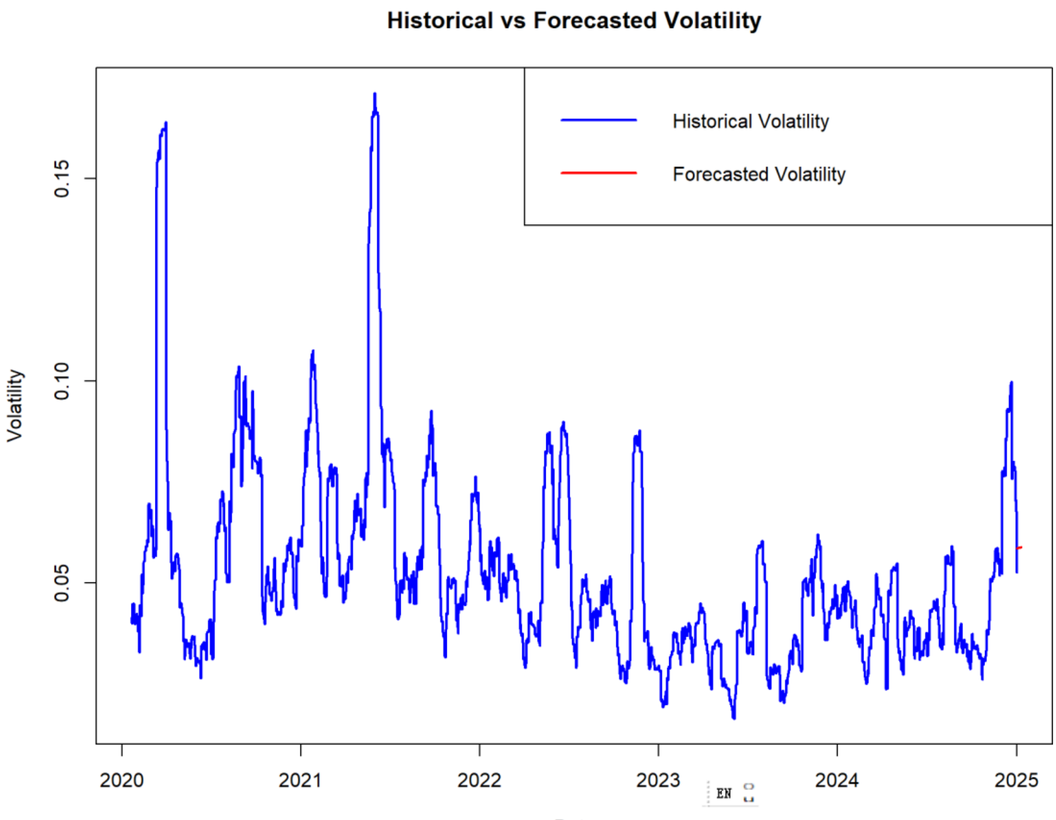
I set 10 days for prediction, and the result shows the volatility in 10 days.

```
Forecasted Mean Returns:
      2025-01-01
T+1  0.001378487
T+2  0.001406348
T+3  0.001426492
T+4  0.001441057
T+5  0.001451587
T+6  0.001459201
T+7  0.001464706
T+8  0.001468686
T+9  0.001471564
T+10 0.001473645

Forecasted Volatility (Standard Deviations):
     2025-01-01
T+1  0.05852022
T+2  0.05855022
T+3  0.05857923
T+4  0.05860729
T+5  0.05863442
T+6  0.05866067
T+7  0.05868605
T+8  0.05871060
T+9  0.05873434
T+10 0.05875730
```



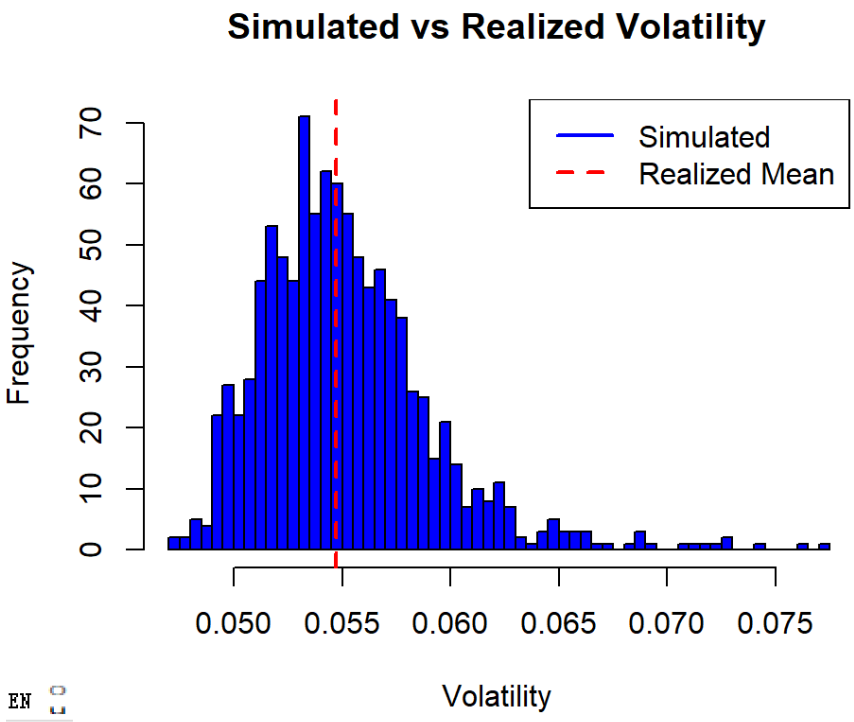
16 By comparing result:

I set history data for rolling window = 20 and do the historical volatility , then I find the image with blue of history data and red is +10 ahead of forcasting and result is in the volatility range



But this is not shown that the result always right with forcasting, so I use monte carlo simulation next steps.

17 .Result :By using monte carlo simulation result:. I set the arma(1,1)garch(1,1) with seed(123) and number of simulation = 1000 (approach by using ugarch)



The graph demonstrates that your GARCH model and Monte Carlo simulation are effective in capturing the **average behavior** of volatility in your dataset. It provides confidence in using the model for forecasting and risk assessment but could benefit from additional analysis to ensure it captures all aspects of real-world volatility dynamic