Regression and Time Series Models

Project 2: Time Series data analysis of Punjab National Bank stock

Submitted by: Chinmay Mutkure

Roll No: 8 A

Submitted to: Dr Amarnath Mitra

Step 1: Check for (Weak) Stationarity:

Augmented Dickey-Fuller (ADF) Test

The Augmented Dickey-Fuller (ADF) test is a statistical test used to determine whether a time series dataset is stationary or not. Stationarity implies that the statistical properties of the series, such as mean and variance, do not change over time.

```
        Augmented Dickey-Fuller Test

data:  stock_price
Dickey-Fuller = -1.8319, Lag order = 13, p-value = 0.6495
alternative hypothesis: stationary
```

Figure 1: Augmented Dickey-Fuller test.

The Augmented Dickey-Fuller (ADF) test was conducted on the stock_price data to check for stationarity. The results are as follows:

Test Statistic (Dickey-Fuller): -1.8319

Lag Order: 13

p-value: 0.6495

Alternative Hypothesis: The time series is stationary

The test statistic value is -1.8319. More negative values indicate stronger evidence against the null hypothesis, which is that the time series has a unit root and is non-stationary. The number of lagged difference terms used in the test regression is 13. This is chosen to ensure that the error term in the regression is white noise.

The p-value is 0.6495. This is the probability of observing the test statistic if the null hypothesis is true. Typically, if the p-value is less than a significance level (like 0.05), we reject the null hypothesis.

However, in this case, the p-value is quite high (0.6495), which means we do not have enough evidence to reject the null hypothesis. Therefore, we cannot infer that the stock_price time series is stationary based on this test.

This implies that the stock price has some trend or seasonality. We might need to apply differencing or other transformations to make it stationary before using it in time series forecasting models. Non-stationary data can lead to misleading statistics and unreliable forecasts. Hence, it's important to make the time series data stationary before using it for further analysis or modeling.

As PNB stock data is non-stationary we Use Transformation (such as First/Second/… Difference | Log | …) to Transform the Data and Check for Stationarity.

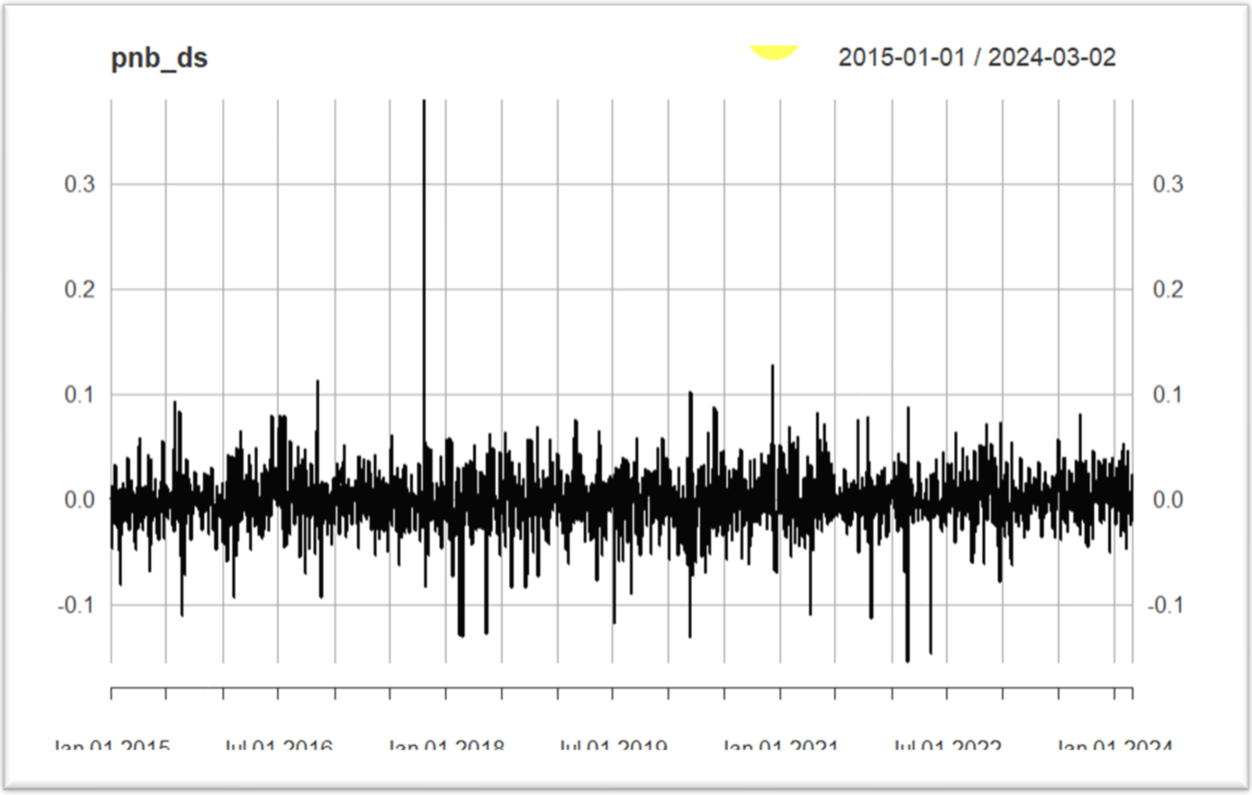

Figure 2 : PNB (First)return Difference Time-Series

Again, we will do the ADF test:

```
Warning: p-value smaller than printed p-value
        Augmented Dickey-Fuller Test

data:  pnb_ds
Dickey-Fuller = -11.874, Lag order = 13, p-value = 0.01
alternative hypothesis: stationary
```

Figure 3: Augmented Dickey-Fuller test after first return

Here we get our stock to be stationary which we will be using for further analysis.

Test Statistic (Dickey-Fuller): -11.874

Lag Order: 13

p-value: 0.01

Alternative Hypothesis: The time series is stationary.

The test statistic value is -11.874. More negative values indicate stronger evidence against the null hypothesis, which is that the time series has a unit root and is non-stationary. The number of lagged difference terms used in the test regression is 13. This is chosen to ensure that the error term in the regression is white noise.

The p-value is 0.01. This is the probability of observing the test statistic if the null hypothesis is true. Typically, if the p-value is less than a significance level (like 0.05), we reject the null hypothesis.

In this case, the p-value is quite low (0.01), which means we have strong evidence to reject the null hypothesis. Therefore, we can infer that the pnb_ds time series is stationary based on this test.

This implies that the stock price does not have any unit root, and it is constant over time with no trend or seasonality. This is a good characteristic for time series forecasting models as they require the data to be stationary.

Step 2: Check for Autocorrelation: Ljung-Box Test

```
        Box-Pierce test

data:  pnb_ds
X-squared = 1.0873, df = 1, p-value = 0.2971
```

Figure 4: Ljung-Box Pierce Test

The Box-Pierce test was conducted on the pnb_ds data to check for autocorrelation in the residuals of a time series model. The results are as follows:

Test Statistic (X-squared): 1.0873

Degrees of Freedom (df): 1

p-value: 0.2971

The test statistic value is 1.0873. This value follows a chi-square distribution with degrees of freedom equal to the number of lags tested. In this case, the degrees of freedom is 1.

The p-value is 0.2971. This is the probability of observing the test statistic if the null hypothesis is true. The null hypothesis of the Box-Pierce test is that the data is independently distributed, meaning there is no autocorrelation. Typically, if the p-value is less than a significance level (like 0.05), we reject the null hypothesis.

However, in this case, the p-value is quite high (0.2971), which means we do not have enough evidence to reject the null hypothesis. Therefore, we can infer that there is no significant autocorrelation in the pnb_ds time series based on this test.

This implies that the residuals from the time series model are independently distributed and the model has adequately captured the underlying correlation structure of the data. This is a good characteristic for time series forecasting models as they require the residuals to be independently distributed.

<u>Step 3:</u> Model for Autocorrelation: ARIMA Models

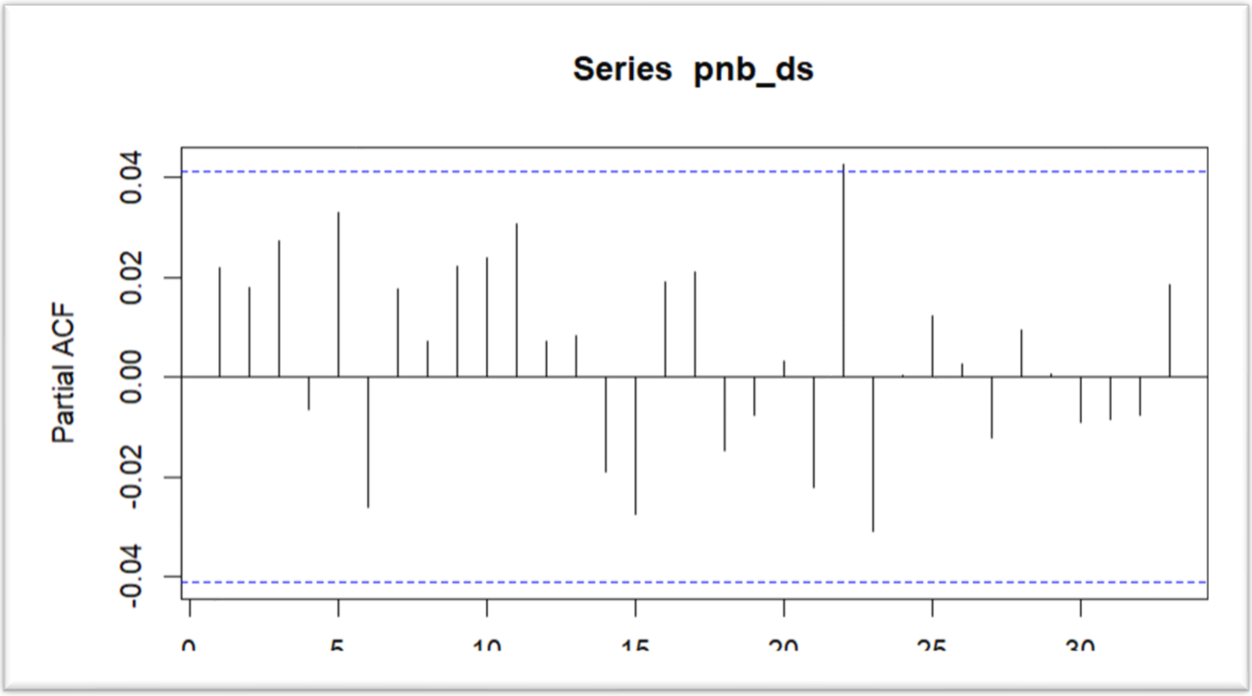

Figure 5: ACF PLOT

```
Series: pnb_ds
ARIMA(1,1,0) with drift

Coefficients:
          ar1   drift
      -0.4982  0e+00
s.e.   0.0182  5e-04

sigma^2 = 0.001106:  log likelihood = 4491.61
AIC=-8977.22   AICc=-8977.2   BIC=-8960.04
```

Figure 6: ARIMA

ARIMA(1,1,0): This is the order of the ARIMA model. The three parameters represent the order of the autoregressive part (AR=1), the degree of differencing (I=1), and the order of the moving average part (MA=0).

Coefficients:

ar1 = -0.4982: This is the coefficient for the autoregressive term. It suggests that the series corrects itself from past values.

drift = 0: This indicates that there is no drift in the model.

s.e.: These are the standard errors of the coefficients. They are used to test the hypothesis that each coefficient is different from zero.

sigma^2 = 0.001106: This is the estimated variance of the residuals (noise) in the model.

log likelihood = 4491.61: This is the log-likelihood, which measures the goodness-of-fit of the model. Higher values are better.

AIC = -8977.22, AICc = -8977.2, BIC = -8960.04: These are information criteria that balance the goodness of fit with the complexity of the model. Lower values are better. The AIC (Akaike Information Criterion) and AICc (corrected AIC) are used for model selection, with the AICc correcting small sample sizes. The BIC (Bayesian Information Criterion) is similar but includes a stronger penalty for models with more parameters.

In summary, the ARIMA(1,1,0) model suggests that the pnb_ds series is a stationary series with one order of differencing and a significant autoregressive component of order 1. The model fits the data well as indicated by the high log-likelihood and low AIC, AICc, and BIC values. However, the model does not include a drift term, suggesting that the series does not have a systematic linear trend.

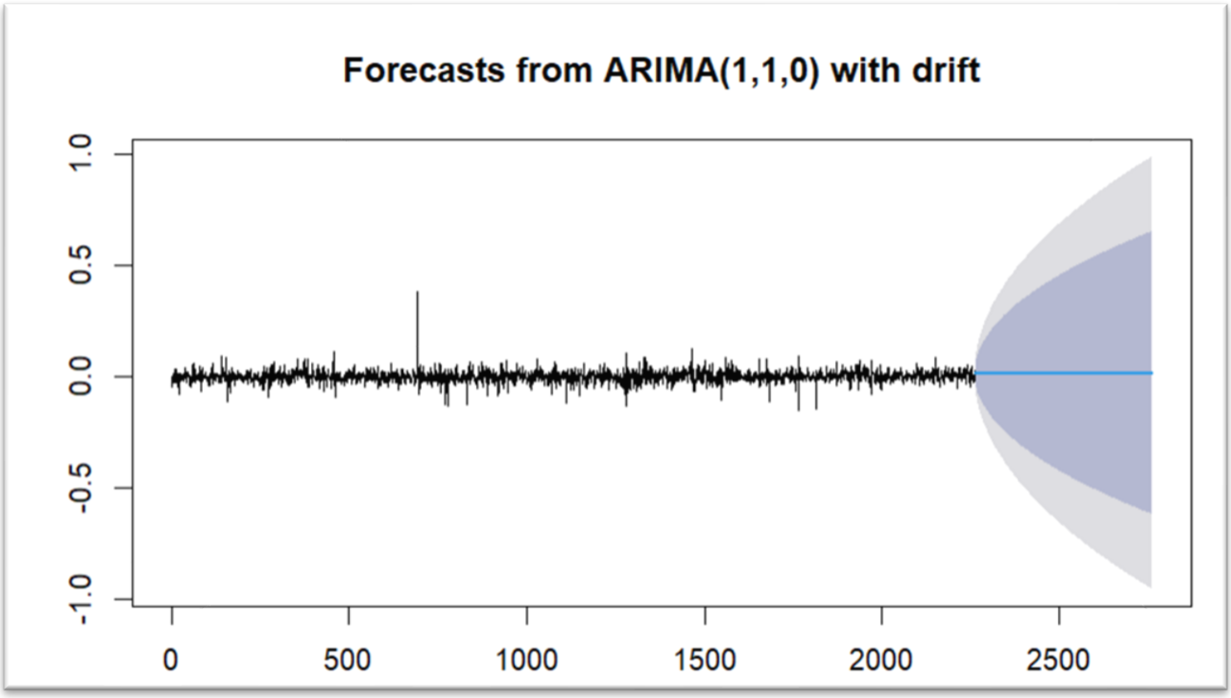

Figure 7: Forecasts from ARIMA (1,1,0) With drift

```
        Box-Pierce test

data:  arma_pq_pnb_ds$residuals
X-squared = 64.74, df = 1, p-value = 8.882e-16
```

Figure 8: Ljung-Box Test for Autocorrelation - Model Residuals

The Box-Pierce test was conducted on the residuals of the arma_pq_pnb_ds model to check for autocorrelation. The results are as follows:

Test Statistic (X-squared): 64.74

Degrees of Freedom (df): 1

p-value: 8.882e-16

The test statistic value is 64.74. This value follows a chi-square distribution with degrees of freedom equal to the number of lags tested. In this case, the degrees of freedom is 1.

The p-value is 8.882e-16. This is the probability of observing the test statistic if the null hypothesis is true. The null hypothesis of the Box-Pierce test is that the data is independently distributed, meaning there is no autocorrelation. Typically, if the p-value is less than a significance level (like 0.05), we reject the null hypothesis.

In this case, the p-value is extremely low (8.882e-16), which means we have strong evidence to reject the null hypothesis. Therefore, we can infer that there is significant autocorrelation in the residuals of the arma_pq_pnb_ds model based on this test.

This implies that the model has not captured all the information in the data, and there is still some pattern in the residuals that the model has not accounted for. This could affect the accuracy of the model's predictions.

This concludes the Box-Pierce Test section of the report. The next steps would involve investigating the source of the autocorrelation and possibly refining the model to better capture the underlying patterns in the data.

<u>Step 4</u>: Check for Heteroskedasticity: ARCH LM Test

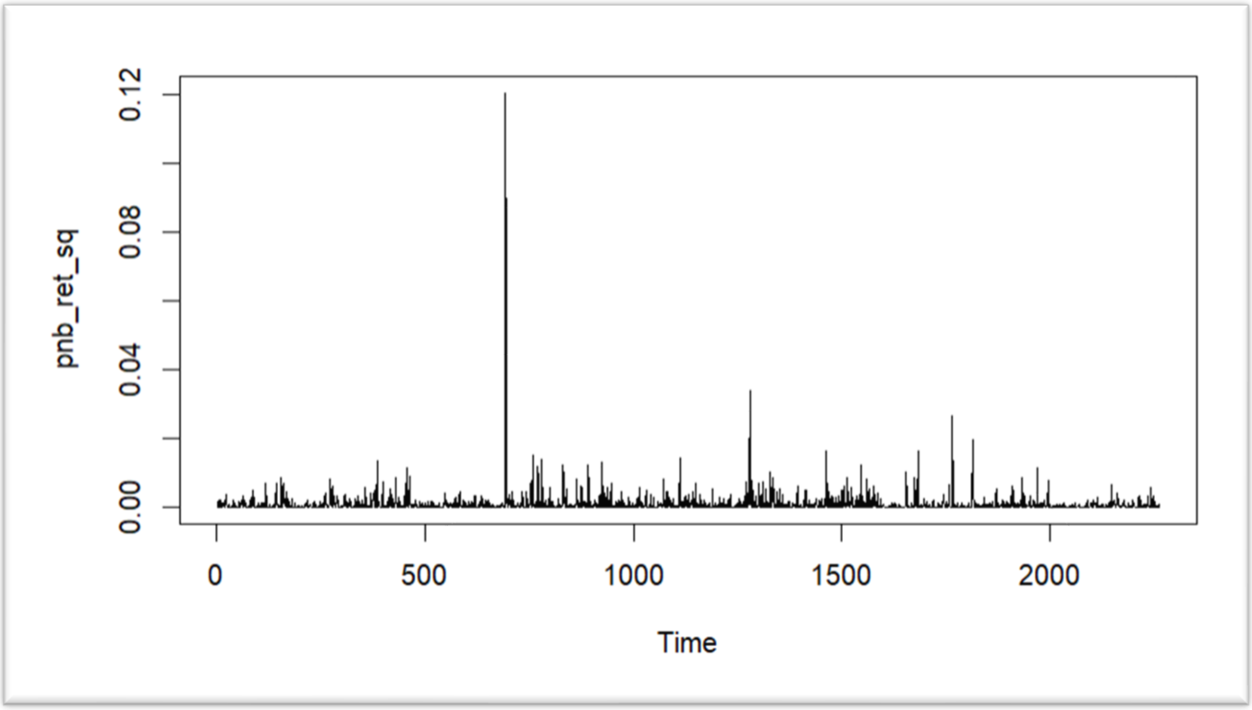

Figure 9 & 10: Test for Volatility Clustering or Heteroskedasticity: Box Test

```
        Box-Pierce test

data:  pnb_ret_sq
X-squared = 413.7, df = 2, p-value < 2.2e-16
```

Figure 10

```
        ARCH LM-test; Null hypothesis: no ARCH effects

data:  arma_pq_pnb_ds$residuals^2
Chi-squared = 397.82, df = 2, p-value < 2.2e-16
```

Figure 11: Test for Volatility Clustering or Heteroskedasticity: ARCH Test

The ARCH (Autoregressive Conditional Heteroskedasticity) Lagrange Multiplier (LM) test was conducted on the squared residuals of the arma_pq_pnb_ds model to check for ARCH effects. The results are as follows:

Test Statistic (Chi-squared): 397.82

Degrees of Freedom (df): 2

p-value: < 2.2e-16

The test statistic value is 397.82. This value follows a chi-square distribution with degrees of freedom equal to the number of lags tested. In this case, the degrees of freedom is 2.

The p-value is less than 2.2e-16. This is the probability of observing the test statistic if the null hypothesis is true. The null hypothesis of the ARCH LM test is that there are no ARCH effects, meaning there is no conditional heteroskedasticity. Typically, if the p-value is less than a significance level (like 0.05), we reject the null hypothesis.

In this case, the p-value is extremely low (< 2.2e-16), which means we have strong evidence to reject the null hypothesis. Therefore, we can infer that there are significant ARCH effects in the residuals of the arma_pq_pnb_ds model based on this test.

This implies that the volatility of the residuals is not constant over time, but depends on past values. This could affect the accuracy of the model's predictions and suggests that an ARCH or GARCH model might be more appropriate for this data.

This concludes the ARCH LM Test section of the report. The next steps would involve investigating the source of the ARCH effects and possibly refining the model to better capture the underlying volatility in the data.

Step 5a: Model for Heteroskedasticity in [Data | Transformed Data] (Step 2) :: GARCH Models

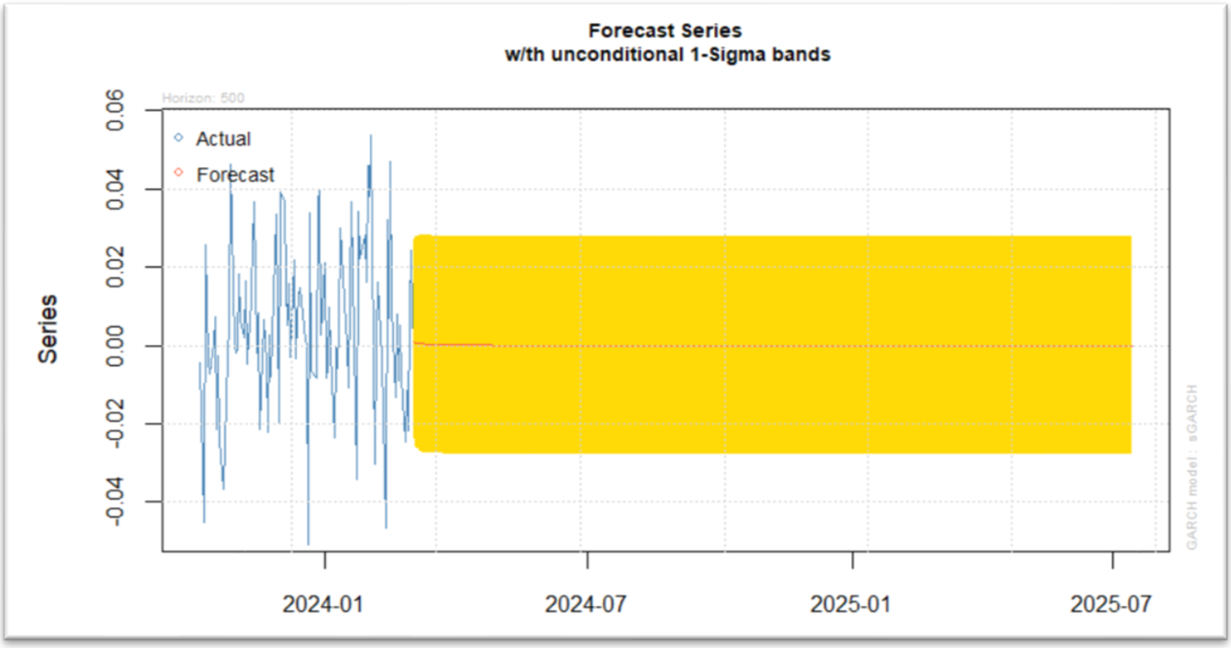

Figure 12 & 13: GARCH forecast plot

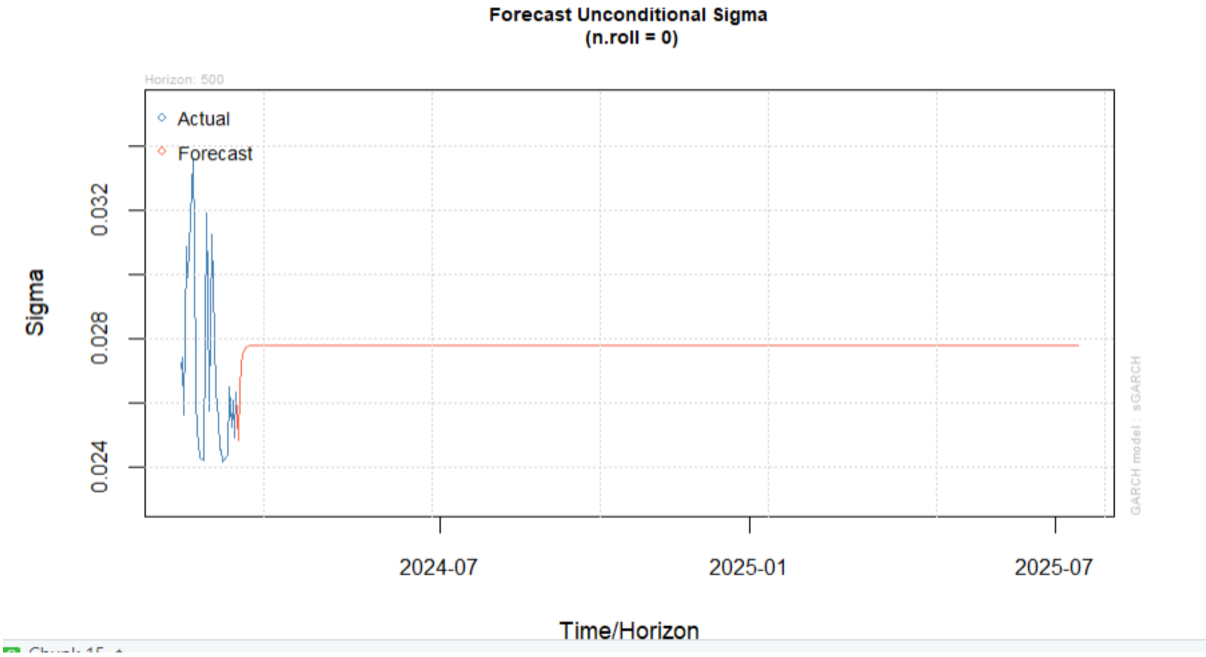
Forecast Unconditional Sigma
(n.roll = 0)
Horizon: 500
Actual
Forecast
Sigma
0.024
0.028
0.032
2024-07
2025-01
2025-07
Time/Horizon
GARCH model : sGARCH

Figure 13