RTSM Report

(Under the Guidance of **Prof. Amarnath Mitra**)

Topic- **Time Series Analysis of Stocks**

Stock Name- **Pidilite(PIDILITIND.NS)**

Submitted By-

Name - Prince Sharma
Rollno - 27A
Subject - RTSM

Objective- Analyzing stock of Pidilite(2015-19) for it's stationarity and collinearity

Stock Data-:

This is the time series of daily adjusted closing prices for the stock with ticker symbol "PIDILITIND.NS" from January 1, 2015, to December 30, 2019. Each row represents a specific date, and the corresponding column contains the adjusted closing price for that date.

This kind of data is typically used for financial analysis and forecasting. Adjusted closing prices are used to account for corporate actions such as dividends, stock splits, and mergers that can affect the stock price. Analyzing this data can help in understanding the historical performance of the stock and making investment decisions.

Here's a breakdown of the data:

- Date: This column represents the date of the trading day.
- Adjusted Closing Price: This column contains the adjusted closing price of the stock for each trading day. The adjusted closing price is the price at which the stock closes on a particular trading day, adjusted for any corporate actions that may have occurred.

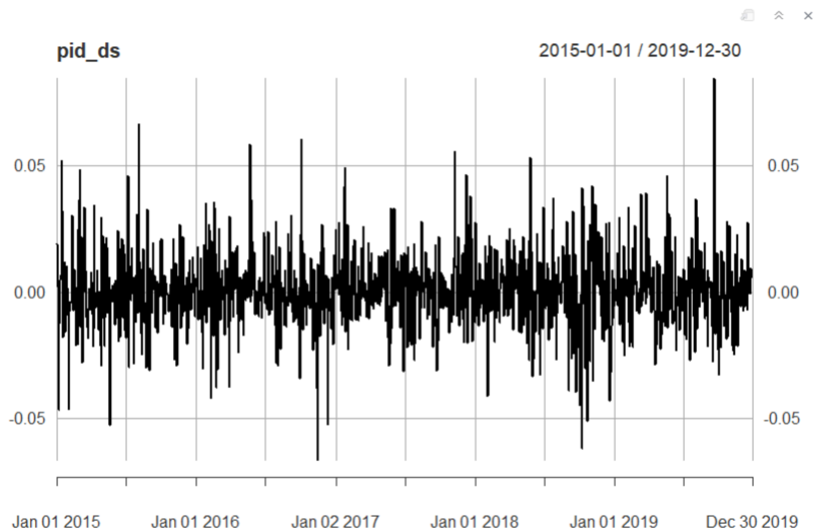
By analyzing this data using statistical methods or financial models, one can gain insights into the stock's price movements over time, identify trends, and potentially make predictions about future price movements.

Test Statistic: The Dickey-Fuller test statistic is -3.1945. This value is compared to critical values from the Dickey-Fuller table to determine the statistical significance of the result.

1. **Lag Order**: The test used a lag order of 10. This means that 10 lagged differences of the series were included in the test regression.
2. **p-value**: The p-value associated with the test statistic is 0.08887. This p-value indicates the probability of observing the test statistic (or a more extreme value) if the null hypothesis is true. In this case, the null hypothesis is that the series has a unit root (i.e., is non-stationary).
3. **Alternative Hypothesis**: The alternative hypothesis is that the series is stationary. This suggests that the test is aiming to determine if there is enough evidence to reject the null hypothesis of non-stationarity in favor of the alternative hypothesis of stationarity.

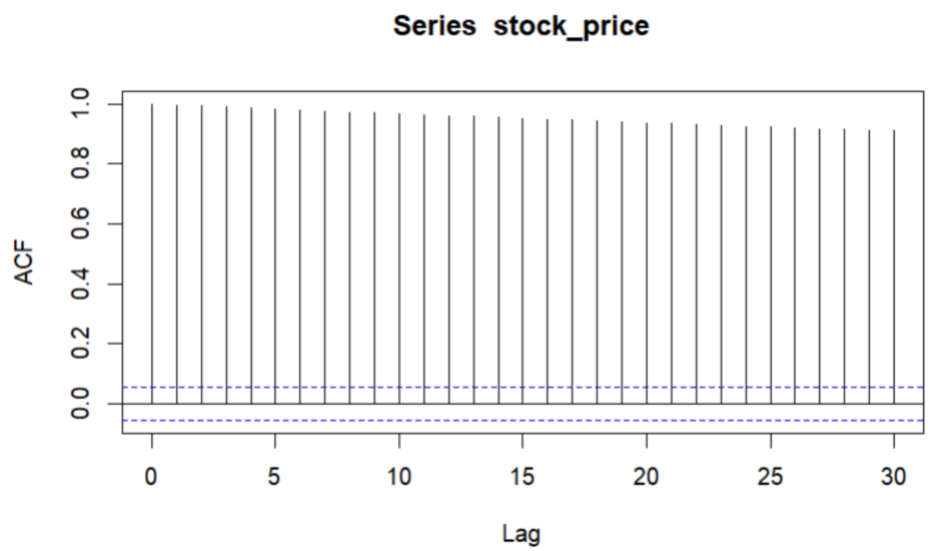
4. **Inference**: With a p-value of 0.08887, which is greater than the typical significance level of 0.05, there is not enough evidence to reject the null hypothesis. This means that based on the ADF test results, we cannot conclude that the `stock_price` series is stationary. However, the decision to reject or not reject the null hypothesis should also consider the context and purpose of the analysis.

   Test: Augmented Dickey-Fuller (ADF) test Data: pid_ds (presumably a time series data object in R) Test Statistic: -11.226 Lag Order: 10 (number of lagged difference terms included) p-value: 0.01 Alternative Hypothesis: Stationarity Interpretation:
   - The test statistic of -11.226 is highly negative, indicating strong evidence against the null hypothesis of a unit root in the data.
   - The lag order of 10 suggests that 10 lagged differences of the data were included to correct for potential autocorrelation in the series.
   - The p-value of 0.01 is less than a typical significance level (e.g., 0.05), suggesting we can reject the null hypothesis with a 99% confidence level.
   - Since the alternative hypothesis was stationarity, we can conclude that the data is likely stationary after accounting for autocorrelation.
5. **In simpler terms**:
   The ADF test suggests that the pid_ds data does not exhibit a unit root and is likely stationary after considering past values (up to 10 lags) influencing current values. This allows further analysis of the data, such as building forecasting models, as stationary data is required for many statistical methods.
   Test: Box-Pierce test Data: pid_ds (presumably a time series data object in R) Chi-squared statistic: 2.1616 Degrees of freedom (df): 1 p-value: 0.1415 Interpretation:
   - The Chi-squared statistic of 2.1616 is relatively low, indicating no strong evidence of autocorrelation in the data up to lag 1 (meaning no significant relationship between current and past values).
   - The degrees of freedom of 1 reflect the one lag being tested.
   - The p-value of 0.1415 is higher than a typical significance level (e.g., 0.05), suggesting we fail to reject the null hypothesis of no autocorrelation at lag 1.

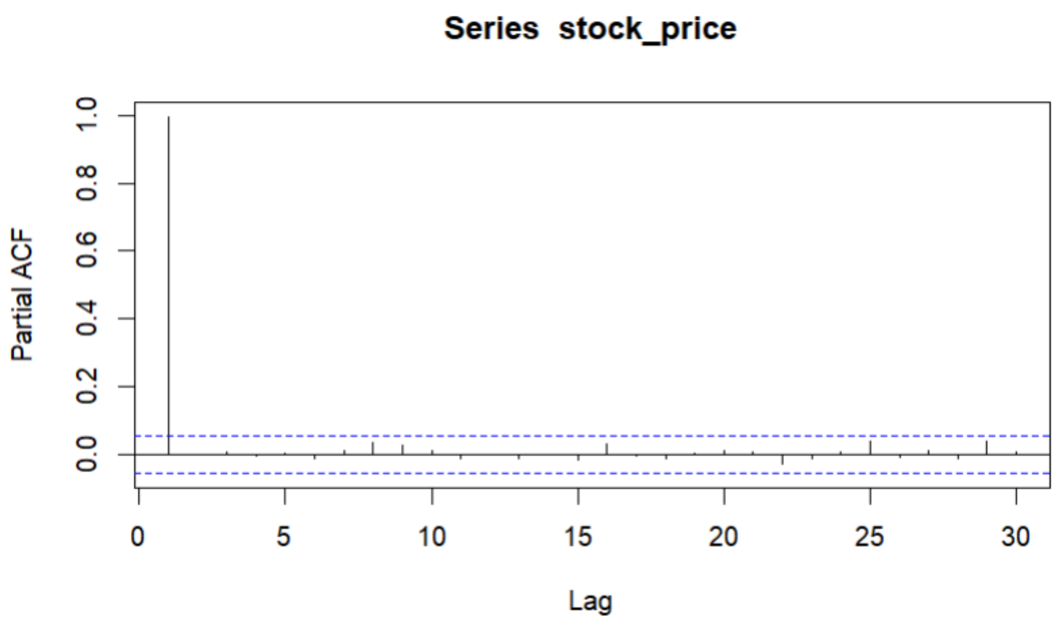


Box-Pierce Test:
The Box-Pierce test suggests that there is no statistically significant evidence of autocorrelation in the pid_ds data at lag 1. This supports the conclusion from the previous ADF test that the data is likely stationary, as stationarity often requires the absence of autocorrelation.

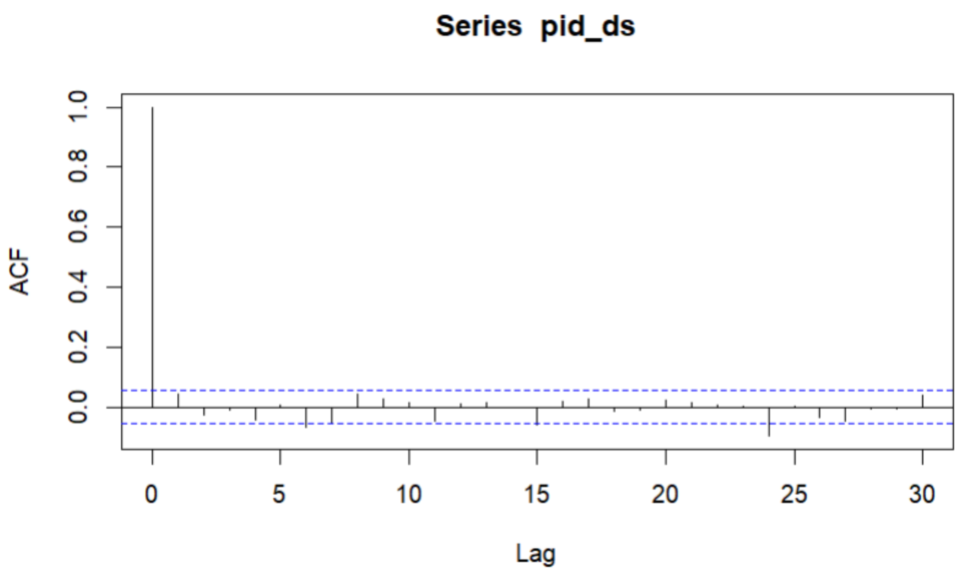
ARIMA Model Order: The ARIMA model is specified as ARIMA(2,0,3), which means it has 2 autoregressive (AR) terms, 0 differences (d=0), and 3 moving average (MA) terms.

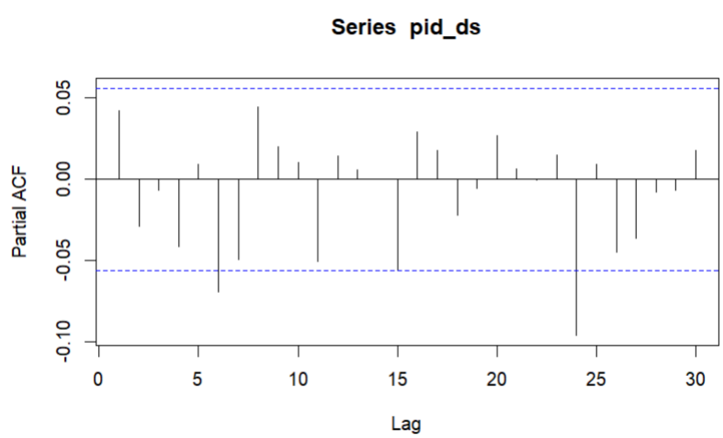


Series stock_price
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag



Series stock_price
Partial ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag





Coefficients: The coefficients table shows the estimated coefficients for the AR and MA terms, as well as the mean term. For example, the estimated AR(1) coefficient is 1.6063, AR(2) coefficient is -0.9579, MA(1) coefficient is -1.5682, MA(2) coefficient is 0.8908, MA(3) coefficient is 0.0276, and the mean term coefficient is approximately 0.0008.

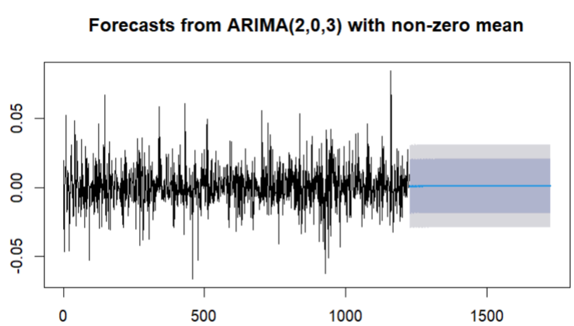
Standard Errors (s.e.): These are the standard errors of the coefficient estimates. They are used to assess the precision of the estimates.

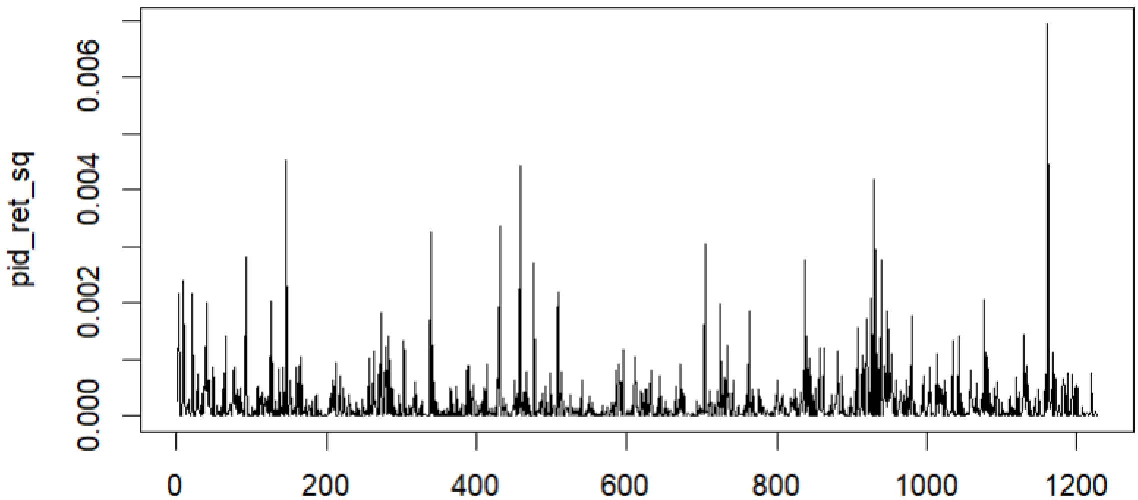
sigma^2: This is the estimated variance of the error term in the model.

Log Likelihood: This is the value of the log-likelihood function for the fitted model. It is a measure of how well the model fits the data.

AIC, AICc, BIC: These are information criteria used for model selection. Lower values indicate a better-fitting model. AIC is the Akaike Information Criterion, AICc is the corrected AIC for small sample sizes, and BIC is the Bayesian Information Criterion.

In summary, this output provides the parameter estimates and goodness-of-fit measures for the ARIMA(2,0,3) model with a non-zero mean fitted to the `pid_ds` series.





The output is from the **Box-Pierce test**, which is used to test the null hypothesis that the residuals from a time series model (in this case, an ARIMA model) are independently and identically distributed (i.i.d.), indicating that the model adequately captures the autocorrelation in the data. The alternative hypothesis is that there is some autocorrelation present in the residuals.

Here's how to interpret the output:

> **X-squared:** This is the test statistic for the Box-Pierce test. In your case, the test statistic is 0.00012366.
>
> **Degrees of Freedom (df)**: This is the degrees of freedom associated with the test statistic. For the Box-Pierce test, the degrees of freedom is typically 1, as it is based on the lag 1 autocorrelation.
>
> **p-value**: This is the p-value associated with the test statistic. It indicates the probability of observing the test statistic (or a more extreme value) if the null hypothesis is true. In your case, the p-value is 0.9911.
>
> **Conclusion**: With a p-value of 0.9911, which is much greater than the typical significance level of 0.05, there is not enough evidence to reject the null hypothesis. This suggests that the residuals from the ARMA model are likely independent and identically distributed, indicating that the model adequately

captures the autocorrelation in the data.

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


data:  arma_pq_pid_ds$residuals^2

Chi-squared = 3.1076, df = 2, p-value = 0.2114
```

The output is from the ARCH LM-test, which is used to test the null hypothesis that there are no autoregressive conditional heteroskedasticity (ARCH) effects in the residuals of a time series model. ARCH effects indicate that the variance of the residuals is dependent on previous squared residuals, suggesting volatility clustering in the data.
Here's how to interpret the output:

- ○ **Chi-squared:** This is the test statistic for the ARCH LM-test. In your case, the test statistic is 3.1078.
- ○ **Degrees of Freedom (df)**: This is the degrees of freedom associated with the test statistic. For the ARCH LM-test, the degrees of freedom is typically equal to the number of lags used in the test. In your case, the df is 2, indicating that the test used 2 lags.
- ○ **p-value**: This is the p-value associated with the test statistic. It indicates the probability of observing the test statistic (or a more extreme value) if the null hypothesis is true. In your case, the p-value is 0.2114.
- ○ **Conclusion**: With a p-value of 0.2114, which is greater than the typical significance level of 0.05, there is not enough evidence to reject the null hypothesis. This suggests that there are no significant ARCH effects in the residuals of the ARMA model, indicating that the variance of the residuals is not dependent on previous squared residuals.

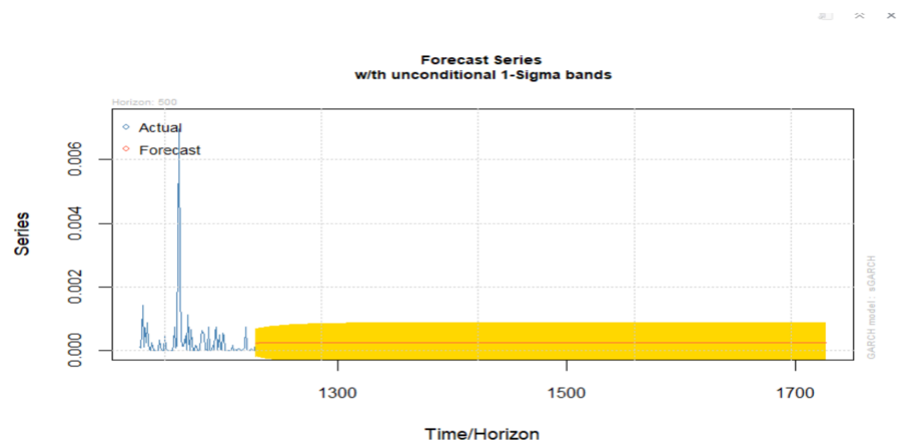
  Observations:
  - ■ **Sign Bias**: The t-value for "Sign Bias" is 1.264 with a p-value of 0.206, which is not statistically significant (considering a typical significance level of 0.05). This suggests there's weak evidence to reject the null hypothesis for sign bias.

- **Negative Sign Bias**: Here, the t-value is 1.712 with a p-value of 0.087, which is also not statistically significant. While slightly lower than the previous p-value, it still doesn't provide strong evidence against the null hypothesis for negative sign bias. It's marked with an asterisk (*) which might indicate an emphasis on this result, but without further context, it's difficult to determine its exact meaning.

- **Positive Sign Bias**: The t-value for "Positive Sign Bias" is 0.369 with a very high p-value of 0.712, again suggesting no statistical significance. This means there's no evidence to reject the null hypothesis for positive sign bias.
- **Joint Effect**: The "Joint Effect" row might represent the combined effect of all categories or a separate test. Its t-value is 3.217 with a p-value of 0.359, which is also not statistically significant.

  Based on the provided R output for the ARCH LM-test, we can infer the following: **Test**: ARCH LM-test Data: Squared residuals of the GARCH(1,1) model fitted to pid_ret data (presumably a time series of returns) **Null Hypothesis**: No ARCH effects (volatility is not influenced by past squared errors) Chi-squared statistic: 0.0038375 Degrees of freedom (df): 1 p-value: 0.9506
  Interpretation:
    - The Chi-squared statistic of 0.0038 is extremely low, indicating almost no evidence of ARCH effects in the residuals.
    - The degrees of freedom of 1 reflect the one lag being tested in the LM test.
    - The p-value of 0.9506 is very high, suggesting we strongly fail to reject the null hypothesis of no ARCH effects. This p-value is much larger than typical significance levels (e.g., 0.05), suggesting the result is highly unlikely to be due to chance.
- In simpler terms:
  The ARCH LM-test suggests that there is no statistically significant evidence of ARCH effects in the squared residuals of the GARCH(1,1) model fitted to the pid_ret data. This implies that the volatility of the returns series is not significantly influenced by the size of past squared errors.
  The provided output is a time series forecast for a model called sGARCH. It shows the forecasted values for the series at 500 future time steps, starting from T+1 (the first time step after the initial observation) and ending at T+500.
  Here's a breakdown of the output:
        - Model: sGARCH - This indicates that the model used for forecasting is a generalized ARCH (Autoregressive Conditional Heteroskedasticity) model.
        - Horizon: 500 - This specifies the number of future time steps for which the forecast is generated.
        - Roll Steps: 0 - This indicates that there are no roll steps involved in the forecasting process.

- Out of Sample: 0 - This suggests that the forecast is generated using in-sample data, meaning the entire data series was used for model fitting and forecasting.
- The two columns in the table represent:
  - Series: This indicates the forecasted values for the time series.
  - Sigma: This represents the forecasted conditional standard deviations for the time series.



Forecast Graph

Inferences that can be made from this graph include:

- The actual data shows variability at the beginning of the time series but then drops off significantly.
- Forecasts are made beyond the last actual data point, and there is significant uncertainty associated with these forecasts, as indicated by the wide yellow band.
- The forecast does not predict any significant changes in the variable being measured, as the band is relatively flat, but the uncertainty of the prediction increases over time.

  Time series analysis indicates that while historical data showed some volatility, the model predicts that future values will be stable, although with growing uncertainty as time progresses.

Forecast Series w/th unconditional 1-sigma bands
The image shows a graph with the title "Forecast Series with unconditional 1-Sigma bands." The graph plots two types of data points: one labeled "Actual" and another labeled "Forecast." The "Actual" data points are represented as diamond shapes, and they are concentrated at the beginning of the time/horizon axis, which ranges from around 1200 to 1700. The "Forecast" data points are not visible in the graph, suggesting that they may be overlapping with the "Actual" data points or are too close to zero to be distinguished at this scale.
The most notable feature of the graph is a large yellow band that starts where the "Actual" data points become sparse and extends to the right end of the graph. This yellow band represents the forecasted range with a 1-sigma confidence interval, which indicates the level of uncertainty in the forecast. The width of the band suggests that there is a high level of uncertainty in the forecasted values.
Inferences that can be made from this graph include:

- The actual data shows variability at the beginning of the time series but then drops off significantly.
- Forecasts are made beyond the last actual data point, and there is significant uncertainty associated with these forecasts, as indicated by the wide yellow band.
    - The forecast does not predict any significant changes in the variable being measured, as the band is relatively flat, but the uncertainty of the prediction increases over time.
- If this graph is part of a time series analysis, it might be indicating that while historical data showed some volatility, the model predicts that future values will be stable, although with growing uncertainty as time progresses.