# REPORT 2

# TOPIC: ANALYSIS AND FORECASTING OF STOCKS WIPRO

# 

# A blue and white logo Description automatically generated

# **MST 21: Regression & Time Series Models**

**Prof. Amarnath Mitra**

**SUBMITTED BY: -**

# MBA (Business Analytics),

# Indian Institute of Foreign Trade, Delhi

**ABSTRACT**

**Time Series Data In case of Wipro:**

The time series analysis of Wipro Ltd stock that we took from yahoo finance provides valuable insights into its market performance over the past few years. It offers investors and analysts an opportunity to understand the trends, patterns, and volatility in Wipro's stock prices, aiding in investment decision-making. Analysing the time series data of Wipro Enterprises Ltd stock offers valuable insights into its market performance, economic significance, industry dynamics, strategic initiatives, regulatory environment, technological innovations, and global macro trends. This information aids investors, analysts, policymakers, and stakeholders in making informed decisions and understanding Wipro's role in the Indian economy and global markets.

**Data collection was done in R code:**

* **Stock Symbol:** WIPRO.NS (Wipro Enterprises Ltd)
* **Date Range:** January 1, 2019, to December 31, 2023
* **Data Retrieval:** Retrieved stock data from Yahoo Finance using the” quantmod “ package.

A computer screen shot of a program

Description automatically generated

**Data Cleaning:**  Removed missing values from the closing prices.

A computer screen shot of a program

Description automatically generated

* **Class:** The next step we have is to analyse the class of the data

**Augmented Dickey-Fuller (ADF) Test:**

The Augmented Dickey-Fuller (ADF) test is a statistical test used to determine whether a unit root is present in a time series dataset. A unit root indicates that a time series is non-stationary, meaning its statistical properties (such as mean and variance) change over time and it lacks a constant mean and/or variance.

Null Hypothesis: The time series is non-stationary.

*H*0​: The time series has a unit root.

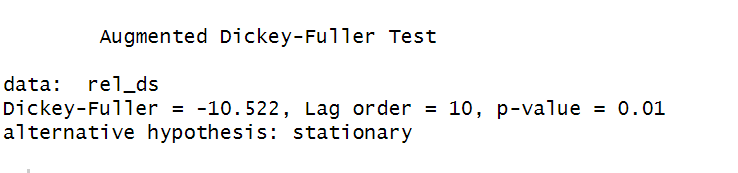
*H*1​: The time series is stationary.

Result: Failed to reject the null hypothesis (p-value = 0.4087).

A screenshot of a computer

Description automatically generated

When we perform the Augmented Dickey Fuller test the p value had been greater than the significant value depicting that we fail to reject the null hypothesis depicting that the data is non-stationary hence we had to **apply log differentiation** to make the data stationary.



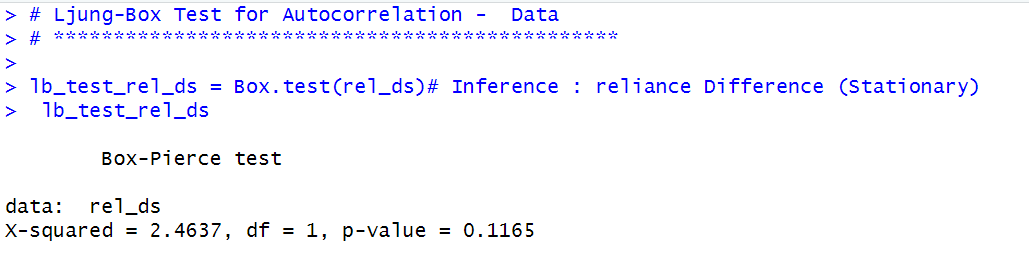


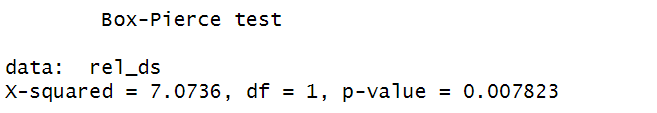
**Autocorrelation and Stationarity:** Time series data often exhibits autocorrelation, where observations at one point in time are correlated with observations at nearby points in time. You'll need to check for and address autocorrelation in your data. Additionally, stationarity (the statistical properties of the data remain constant over time) is an important concept to understand and ensure for accurate modelling.

A white background with black text

Description automatically generated

**The Ljung-Box test**: It is a statistical test used to determine whether there is significant autocorrelation in a time series dataset. Autocorrelation refers to the correlation between observations at different time points within the same series.

 A black text on a white background

Description automatically generated

**Result:**

In this case, since the p-value (0.007823) is much smaller than a typical significance level like 0.05 therefore we reject the null hypothesis. Hence, we have evidence to suggest that there is significant autocorrelation in the time series.

*H*0​: No Auto correlation

*H*1​: Auto Correlation

**Result:** Failed to reject null hypothesis therefore auto corelation exists in the data.

**Autocorrelation Function (ACF):** The ACF measures the correlation between a time series and its lagged values at different time lags. It helps to assess the presence and strength of autocorrelation at various lags. The ACF is plotted as a function of the lag, with the lag on the horizontal axis and the autocorrelation coefficient on the vertical axis.

A graph with lines and numbers

Description automatically generatedA graph with numbers and lines

Description automatically generated

**PACF:** The PACF stands for Partial AutoCorrelation Function. It's a statistical tool used in time series analysis to identify the direct relationship between observations at different time lags while controlling for the effects of other lags. In simpler terms, it measures the correlation between a time series and its lagged values after removing the correlations explained by the intervening lags

A graph with numbers and lines

Description automatically generated

A graph with lines and numbers

Description automatically generated

A graph with numbers and lines

Description automatically generated

A graph with numbers and lines

Description automatically generated

As we can observe that there is downward trend it suggests that the series does not exhibit long-term dependencies and may be suitable for modelling using simpler forecasting methods like ARIMA (AutoRegressive Integrated Moving Average). Hence the data is good for ARIMA testing.

A computer code with black text

Description automatically generatedA black text on a white background

Description automatically generated

In summary, the ARIMA (0,0,2) model with zero mean suggests that the time series can be adequately described by a moving average model without the need for autoregressive terms or differencing. The estimated coefficients provide insight into how past errors influence the current value of the series. It includes two moving average terms (MA), indicating that the current value of the series is influenced by the past two forecast errors (residuals)

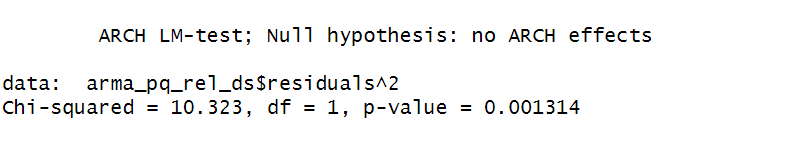
A screen shot of a graph

Description automatically generated

**Moving on for the Box- Pierce test:**   
As we can observe that the p value is less that means that there exists auto corelation between the residuals. A low p-value from the test suggests that there is significant autocorrelation in the residuals, indicating a lack of fit of the model to the data. On the other hand, a high p-value suggests that the residuals are consistent with being independently distributed, indicating a good fit of the model.

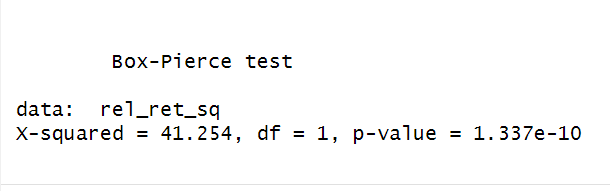
Hence, we proceed further and reject the null hypothesis and conclude that auto corelation exists.

A white background with black text

Description automatically generated

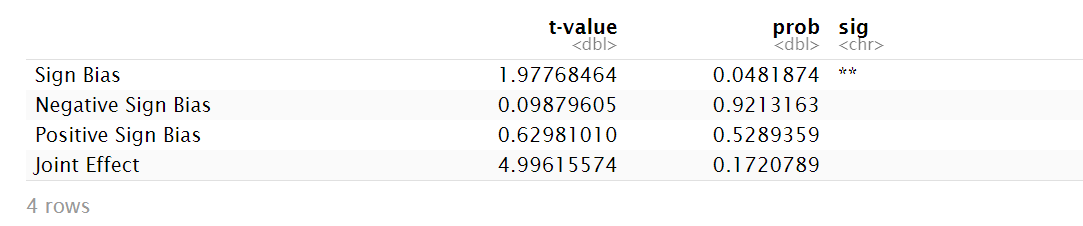
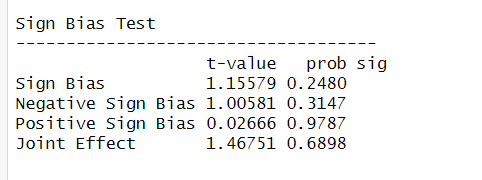
**Test for Volatility Clustering or Heteroskedasticity: Box Test**

It’s used in time series analysis to detect the presence of conditional heteroskedasticity in the residuals of a model. Conditional heteroskedasticity refers to the situation where the variability of the residuals (or errors) in a time series model is not constant over time but instead exhibits clustering or patterns of volatility.



**GARCH FORECAST:**

The GARCH (Generalized Autoregressive Conditional Heteroskedasticity) test, often referred to as the ARCH-LM test, is used to assess whether a time series exhibits autoregressive conditional heteroskedasticity (ARCH) or generalized autoregressive conditional heteroskedasticity (GARCH) effects.



Overall, the output suggests that while there is evidence of overall sign bias in the residuals of the GARCH model, there is no significant negative or positive sign bias, and the joint effect of sign bias is not significant.

A graph with a yellow line

Description automatically generated

A graph of a graph

Description automatically generated

1. **Blue Line (Actual Values)**:
   * Downward trend: The blue line represents a historical or observed downward trend in the variable over time. This could indicate that the variable has been decreasing or experiencing a decline in values over the observed period.
2. **Red Line (Forecasted Values)**:
   * High values: The red line represents the forecasted or predicted values of the variable. If the red line is high while the blue line is declining, it suggests that the model is forecasting an increase or higher values for the variable in the future, despite the current downward trend observed in the data.

In this scenario, the model's forecasts diverge from the current trend observed in the data, indicating that the model predicts a reversal or change in the trend towards higher values. This could imply that the model anticipates a shift in the underlying factors influencing the variable, leading to an upward movement in values despite the historical downward trend.