Predictive Analytics

Assignment 2

Submitted to

Prof. Vamsidhar Ambatipudi

[Department of Management]

Date: 06-12-2022

Submitted By:

Vishva Bhalodiya [2021H1540833P]

Problem Statement: Take any one security and apply different price forecasting mechanisms to the security and identify an appropriate mechanism based on RMSE

Introduction: In this assignment, Different Time series models have been built, and compared the RMSE value for each.

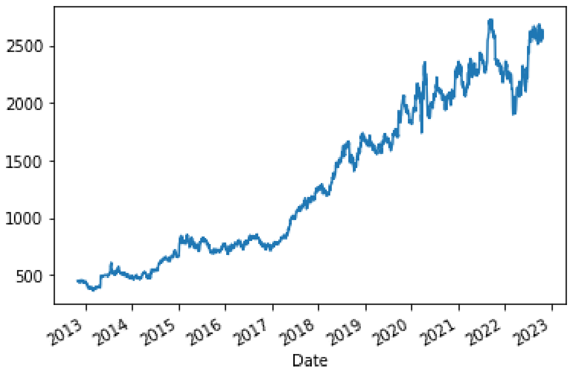
Data Used: The stock "Hindustan Unilever Limited", NSE symbol HINDUNILVR.NS

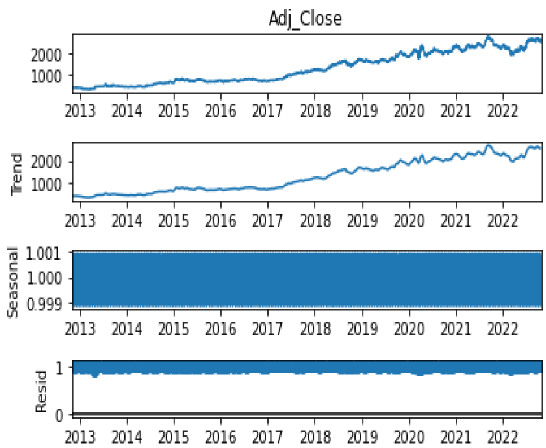
Approach:

1. Identified security (Stock) of Hindustan Unilever Limited (HINDUNILVR.NS)
2. Downloaded the daily price of stock data "HINDUNILVR.NS" from yahoo finance from the period of 1st November 2012 to 31st October 2022
3. Before implementing the model decomposing the data, checking the autocorrelation, and stationarity, differentiating the data
4. Implementation of classic time series models 1) AR, 2) MA, 3) ARIMA, 4) SARIMA, 5) SARIMAX, 6) Simple Exponential Smoothing (SES), 7) VAR, 8) VARMAX
5. Calculate the RMSE value of each classical time series model
6. Choose the appropriate time series model based on the RMSE value

Time Series Decomposition:

The first property of time series is that the timestamp used to identify the data has intrinsic value. In univariate time series models for forecasting, the only variables utilised are the goal variable and its temporal fluctuation. Univariate models are suitable for time series.





Time Series Decomposition is a technique to extract multiple types of variation from your dataset. There are three important components in the temporal data of a time series: seasonality, trend, and noise.

- Seasonality is a recurring movement that is present in your time series variable. For example, the temperature of a place will be higher in the summer months and lower in the winter months. We could compute average monthly temperatures and use this seasonality to forecast future values.
- A trend can be a long-term upward or downward pattern. For example, on top of the summer/winter seasonality, we may well see a slight increase in average temperatures over time.
- Noise is the part of the variability in a time series that can neither be explained by seasonality nor by a trend. When building models, you end up combining different components into a mathematical formula. Two parts of such a formula can be seasonality and trend. A model that combines both will never represent temperature values perfectly; an error will always remain. This is represented by the noise factor.

Inference:
The decomposition of the data shows an upward trend and a strong seasonality.

Stationarity

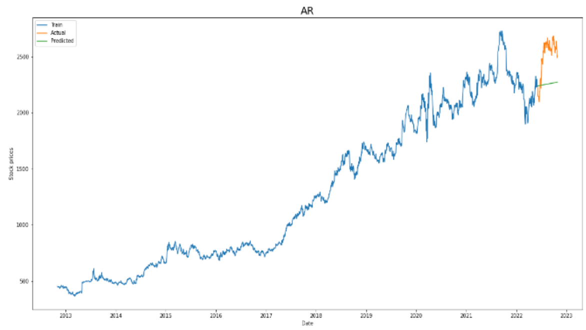
Stationarity is a key component of the definition of a time series. A time series without a trend is said to be stagnant. Some time series models can't account for patterns (more on this later). The Dickey-Fuller Test can be used to identify non-stationarity, and differencing can be used to get rid of non-stationarity.

Observations of the Dickey-fuller test	
Test Statistic	-0.037437
p-value	0.955324
#lags used	24.000000
number of observations used	2438.000000
critical value (1%)	-3.433035
critical value (5%)	-2.862726
critical value (10%)	-2.567401
dtype: float64	

Inference: The presence of a unit root in the time series is the null hypothesis for the ADF test. The data may be stationary, which is an alternate theory. The p-value is the second value. We can reject the null hypothesis (reject non-stationarity) and accept the alternative hypothesis if this p-value is less than 0.05. (stationarity). Because we cannot rule out the null hypothesis in this situation, we must conclude that the data are not stationary.

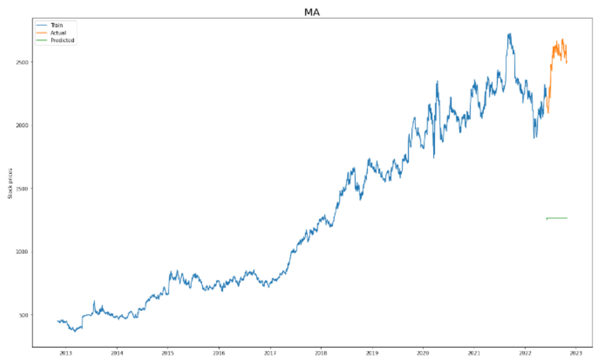
Time Series Model:

1. Autoregression (AR)



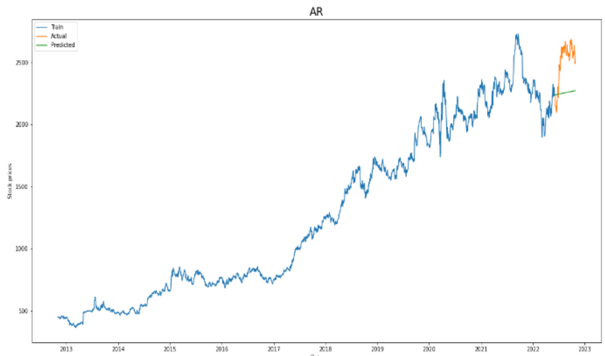
The foundation of the SARIMAX series is autoregression. The AR model can be thought of as a regression model that uses a variable's previous (lagged) values to predict its future value. The number of lagged values to include in the model is indicated by the letter "p" in the order of an AR model. The simplest model is the AR(1) one, which predicts the current value solely based on the previous timestep's value. The entire length of the time series is the most values you are permitted to utilize.
RMSE : 289.9988170097882

2. Moving average (MA)



The second component of the larger SARIMAX model is the Moving Average. Similar to the AR model, it predicts the current value of the variable using data from the past. The Moving Average model does not use the variable's actual values; instead, it uses historical values. Instead, the Moving Average makes predictions based on the forecast inaccuracy of earlier time steps.
RMSE: 1242.284689648154

3. Autoregressive integrated moving average (ARIMA)

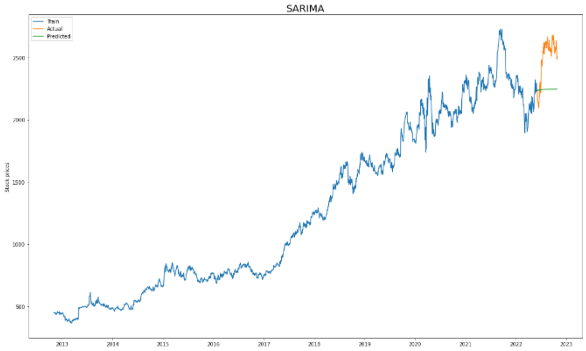


A stationary time series is required by the ARMA model. A time series is stable if it is stationary. If your time series is not stable, we can use the Augmented Dickey-Fuller test to determine this and then use differencing. The ARIMA model enhances the ARMA model with automatic differencing.

The additional argument allows you to specify how many times the time series must be differenced.For example, an ARMA(1,1) that needs to be differenced one time would result in the following notation: ARIMA(1, 1, 1). The first 1 is for the AR order, the second one is for the differencing, and the third 1 is for the MA order. ARIMA(1, 0, 1) would be the same as ARMA(1, 1).Here the order of ARIMA is ARIMA (0,2,1)

RMSE: 268.4225395906145

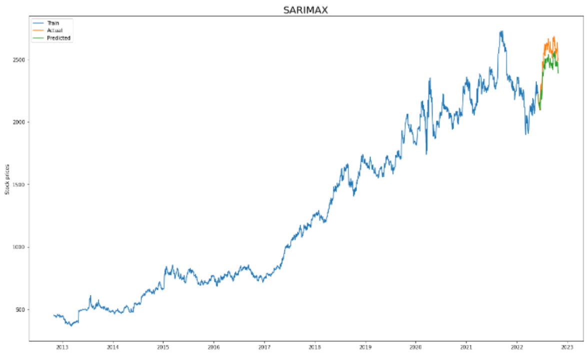
4. Seasonal autoregressive integrated moving-average (SARIMA)



The ARIMA model includes seasonal effects thanks to SARIMA. It is crucial to incorporate seasonality into your forecast if there is any in your time series. Due to the addition of a seasonal parameter to each component's existing regular parameter, SARIMA notation is considerably more complicated than ARIMA.. For example, let's consider the ARIMA(p, d, q) as seen before. In SARIMA notation, this becomes SARIMA(p, d, q)(P, D, Q)m. Here the order of SARI MA is SARIMA order (2, 1, 2), seasonal_order (1, 0, 1, 5)

RMSE: 301.0134662562823

5. Seasonal autoregressive integrated moving-average with Exogenous Regressors (SARIMAX).



The SARIMAX model is the most complicated variation. It combines seasonal effects, AR, MA, and differencing. Additionally, it includes the X: external variables. You might use SARIMAX to include any variables you have that could enhance your model. Here the order of SARIMAX is SARIMAX order (2, 1, 2)
RMSE: 103.65129586894555

6. Simple exponential smoothing (SES)

An adaptation of this simple moving average is exponential smoothing. It uses a weighted average of previous data as opposed to the average. A more current number will count more, whereas a value from the past will count less.
RMSE: 308.8989435397426

Conclusion: In this assignment, our objective is to find the appropriate price forecasting model for stock. After analyzing the different univariate time series models for this stock, we can conclude that the SARIMAX (2,1,2) model gives the lowest RMSE value among other all, but the value of RMSE is still very high, hence none of the above models could give appropriate forecasting. The RMSE can be further reduced by changing the order of the model.