Modeling an Indian Stock Market Company's Stock: Bharat Petroleum Corporation Ltd.

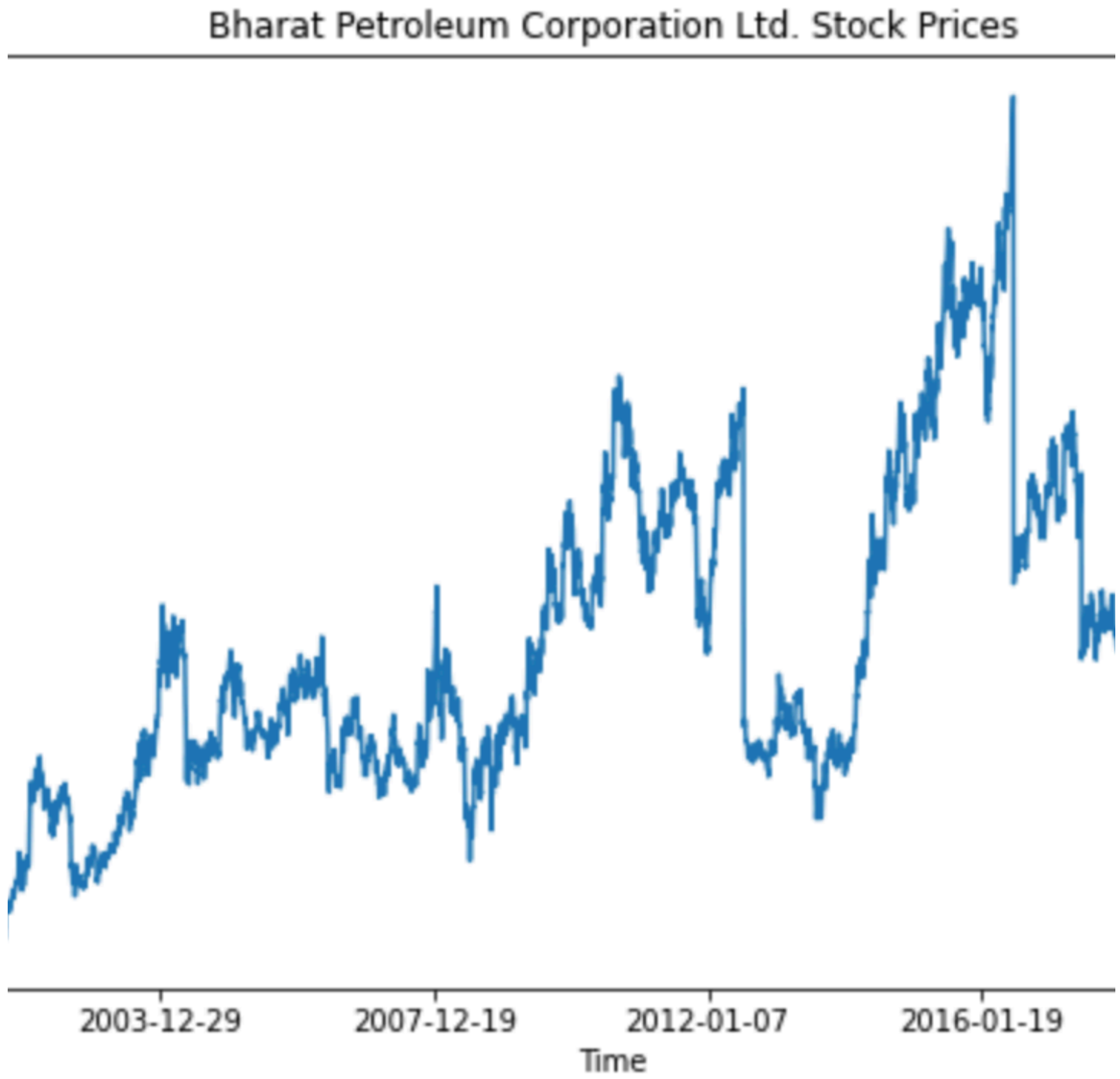
Capstone III Project Report

Elinor Velasquez

Modeling the Bharat Petroleum Corporation Ltd. stock

Capstone III Project

- BPCL is an energy company
- Data is collected over 20 years
- Over time, the stock appears to be slowly increasing (trend)
- This data is time series data of the VWAP (Volume Weighted Average Price)



Hypothesis Test

Capstone III Project

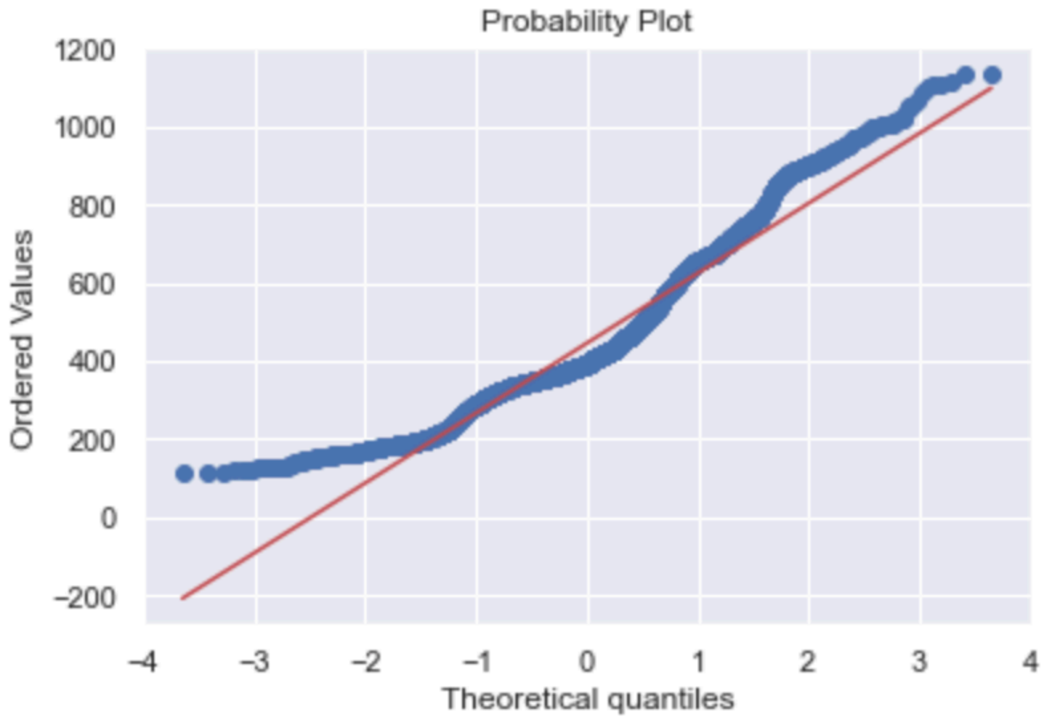
- Null Hypothesis: There is no relationship between VWAP (Volume Weighted Average Price) and other features such as time.
- Alternative Hypothesis: There is some relationship between VWAP and other features such as time.

H_0: $b_1 = 0$
H_A: $b_1 = 0$ is false: $VWAP = b_1 f(X) + e$

If $b_1 = 0$, then $VWAP = b_0 + e$,
e = error,
meaning VWAP has no relation to time (X_1) or any other feature (X_2, X_3, ...)

The distribution of the Volume Weighted Average Price (VWAP)

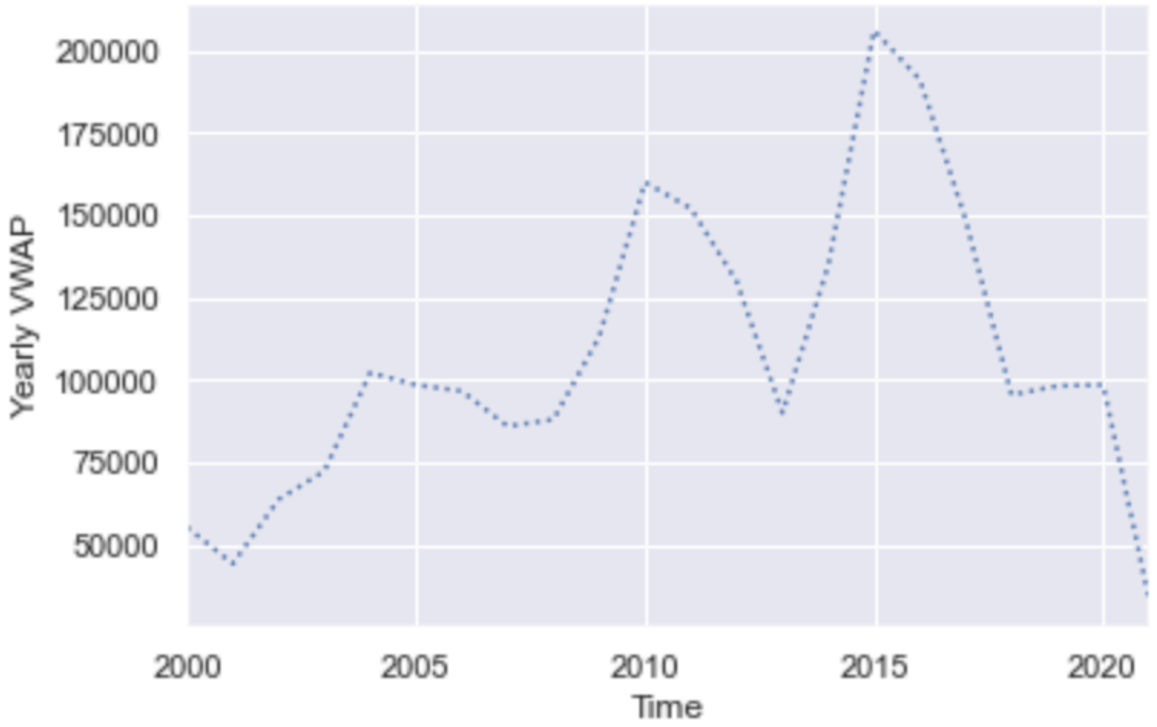
- Q-Q probability plot tests if the VWAP data is normally distributed
- The data is weakly normally distributed according to the Q-Q plot
- Thus, when modeling the VWAP, we cannot normalize the data: We can only standardize the data
- This plotted data is time series data of the VWAP (Volume Weighted Average Price)



The Volume Weighted Average Price of the stock is non-stationary

Capstone III Project

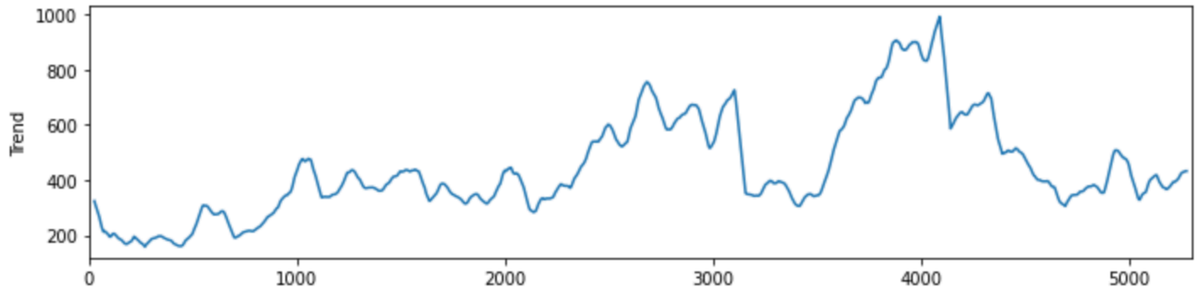
- Resampling the data by year creates a smooth graph
- The plotted points demonstrate an upward trend, with an increasing mean and increasing variance, so the VWAP is non-stationary

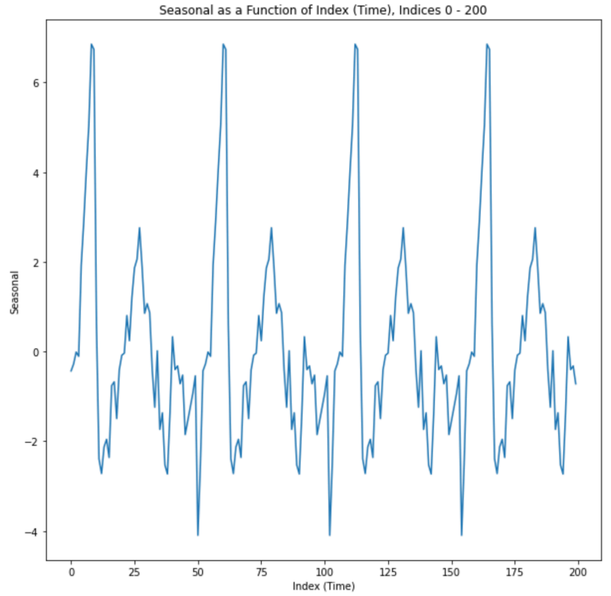


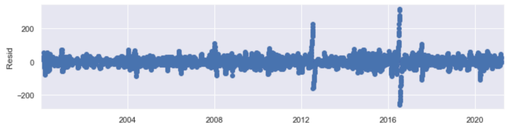
Decomposition of VWAP data shows component behavior

Capstone III Project

- Top Graph: Trend - Slowly increasing mean and increasing variance
- Bottom Left Graph: Seasonality - Highly periodic (Fourier behavior)
- Bottom Right graph: Residual - Additive decomposition of data shows residual component is centered at zero







Model I: Vector AutoRegression

Capstone III Project

- Lag 1 model (t-1 dependence) shows VWAP (Y_1) is dependent on past behaviors of itself and another feature (Y_2)
- Vectorized model allows for the modeling of another feature (Y2), not just the VWAP modeling.
- Mean absolute percentage error (MAPE) score of VAR(5)= 1.15

$$Y_{1,t} = \alpha_1 + \beta_{11,1} Y_{1,t-1} + \beta_{12,1} Y_{2,t-1} + \epsilon_{1,t}$$
$$Y_{2,t} = \alpha_2 + \beta_{21,1} Y_{1,t-1} + \beta_{22,1} Y_{2,t-1} + \epsilon_{2,t}$$

Model II: The Facebook Prophet model

Capstone III project

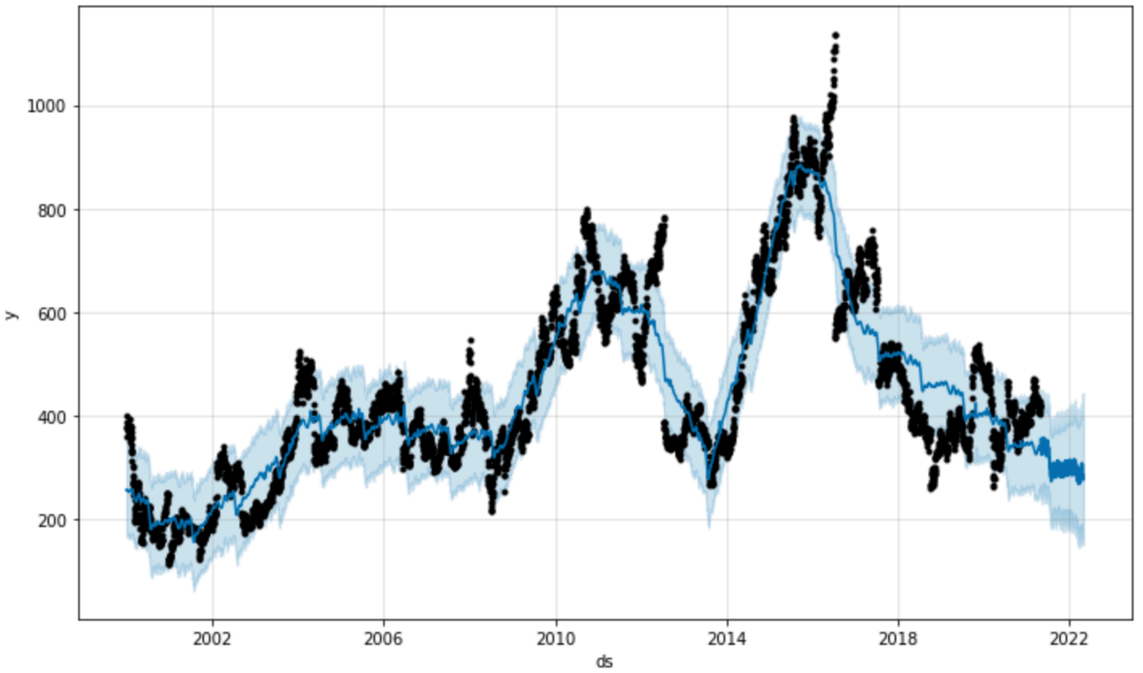
- y(t) is the target (VWAP stock price)
- g(t) is the trend function
- s(t) is the seasonality function
- h(t) is the effects of holidays
- e(t) is the error term
- MAPE score of Prophet = 0.24

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

Model II: The Facebook Prophet model

Capstone III Project

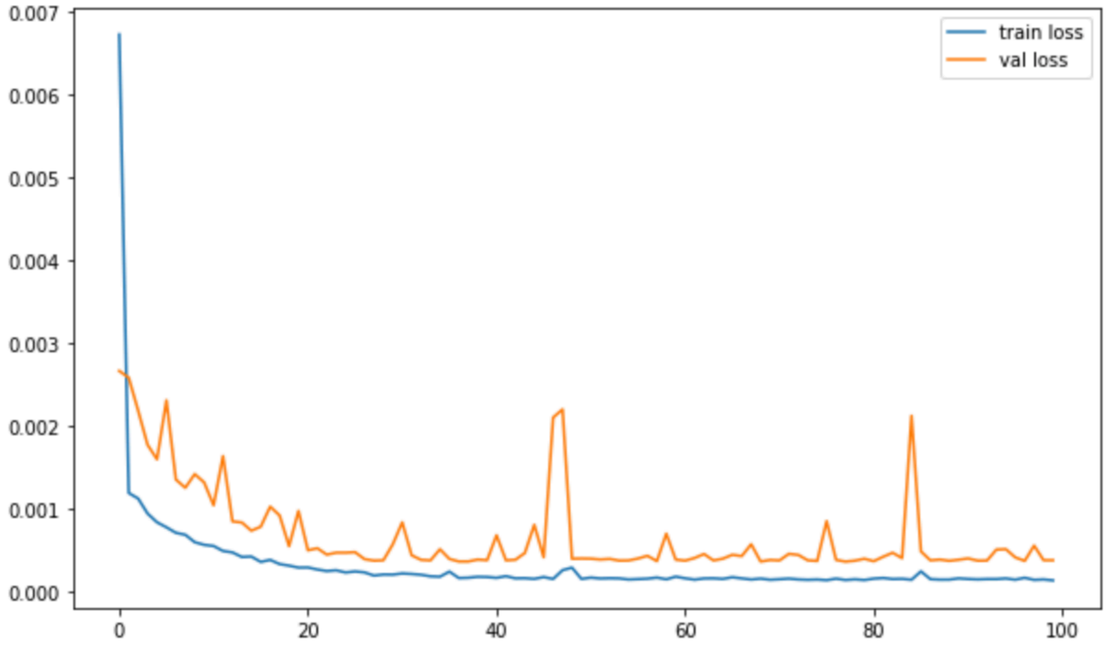
- The vertical axis, y, is the target (VWAP stock price)
- The horizontal axis, ds, is time
- The (blue) line equals the forecasted values of VWAP by Prophet
- The (black) dots are the data points
- The (light blue) shading equals the confidence intervals for the forecasted points



Model III: The Stacked Long Short Term Memory (LSTM) neural network model

Capstone III Project

- The loss function for the Stacked LSTM model was chosen as the mean squared error (MSE)
- The bottom (blue) line is the training data error
- The top (orange) line is the validation (testing) data error
- The horizontal axis equals the number of epochs (number of times the algorithm has been applied to the entire dataset)
- The vertical axis is the MSE



Model III: The Stacked Long Short Term Memory (LSTM) neural network model

Capstone III Project

- The vertical axis is the target (VWAP price)
- The horizontal axis is time (days)
- The (blue) curve spanning 0 to 5000 days equals the target
- The (orange) curve spanning 3500 to 5500 equals the forecasted values
- The stacked LSTM model MAPE score is 0.35.

Future Work

Capstone III Project

- Tune the hyper-parameters with Auto-Keras to improve the Stacked LSTM model
- For the VAR(5) model: Use more features in the model to decrease the MAPE error score
- Add more hidden layers to the stacked LSTM model to improve the model output (Auto-Keras does this automatically)