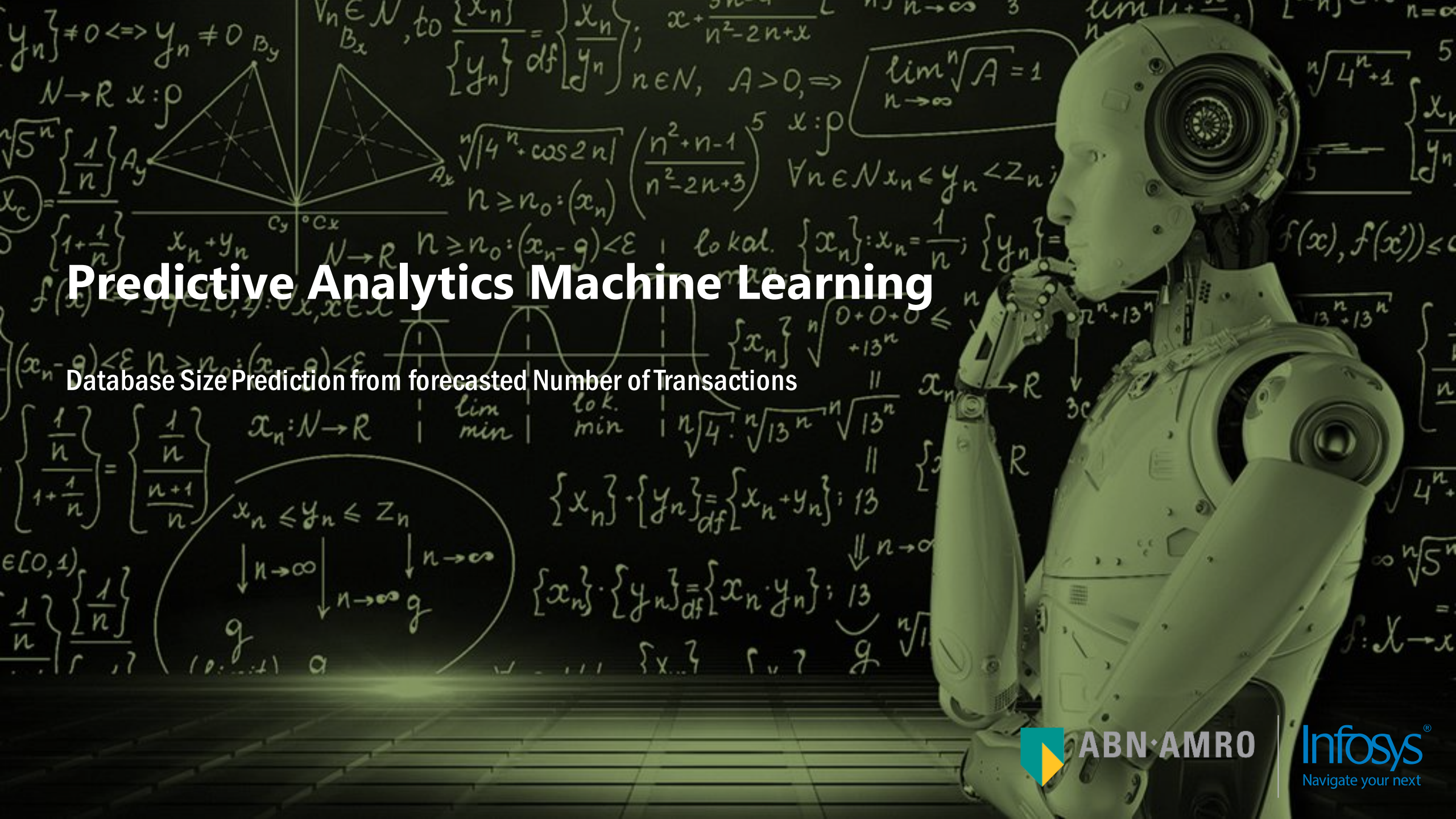
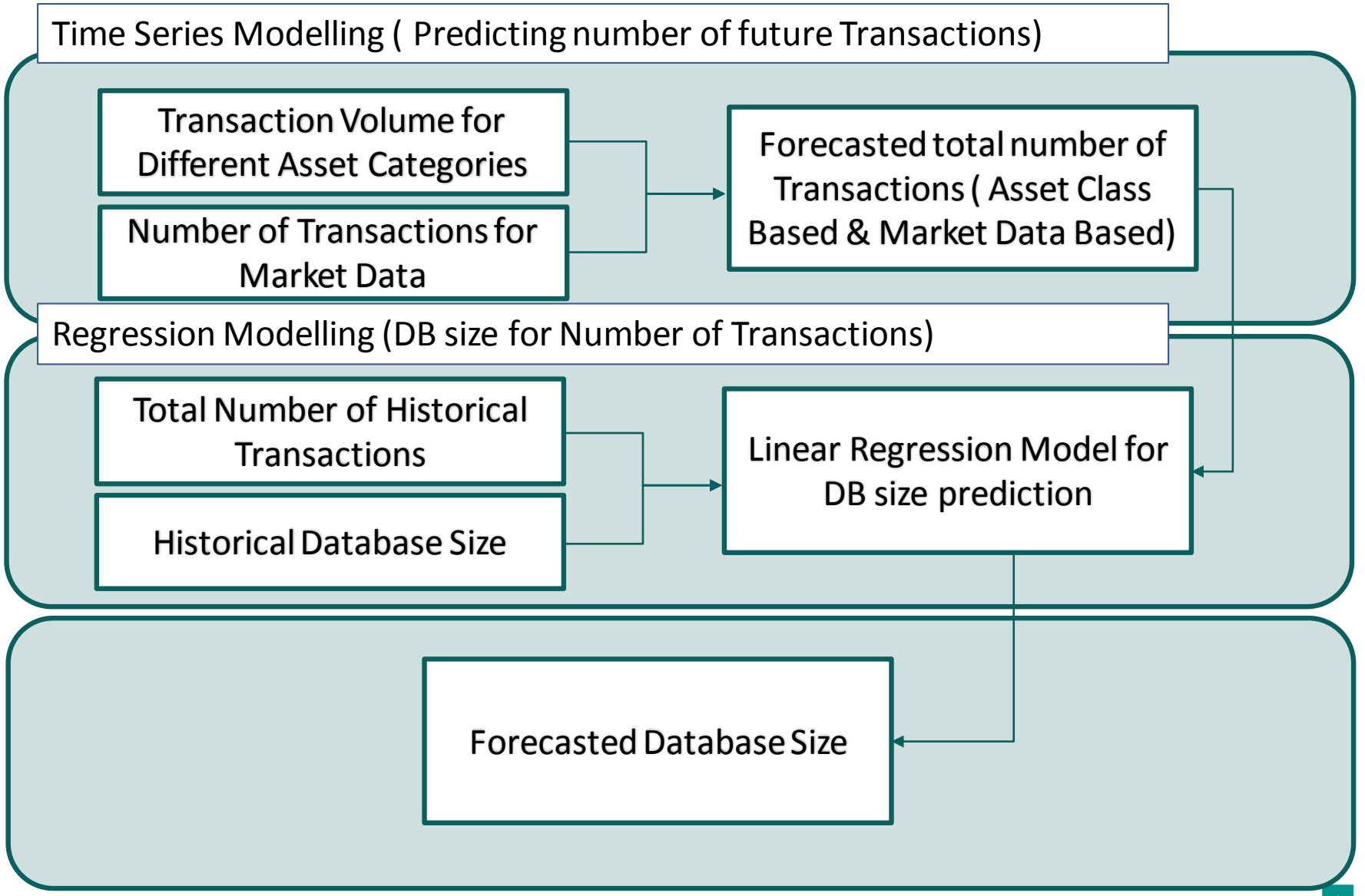
Predictive Analytics Machine Learning

Database Size Prediction from forecasted Number of Transactions



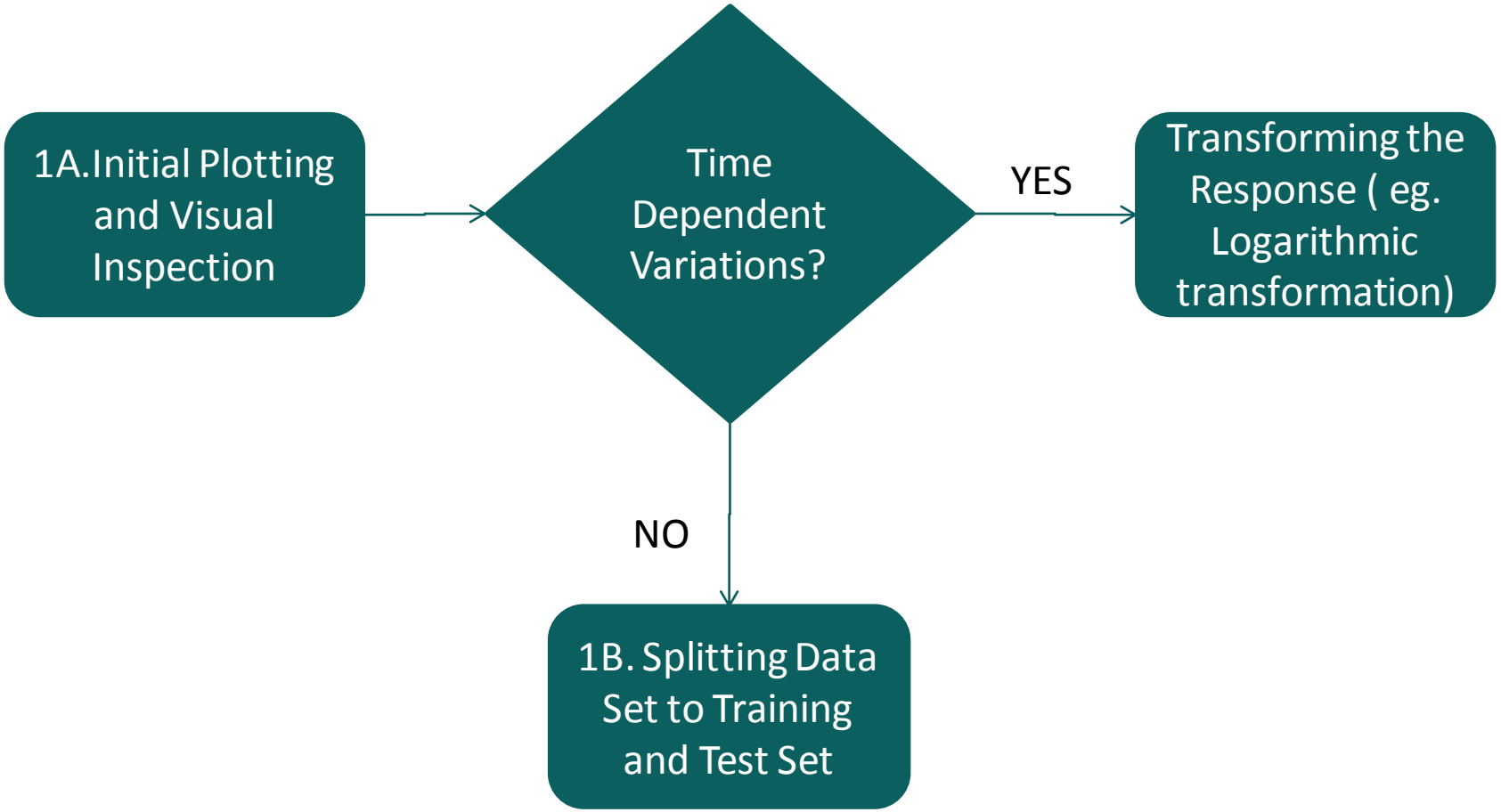
ABN·AMRO | Infosys
Navigate your next

Overall Process Design



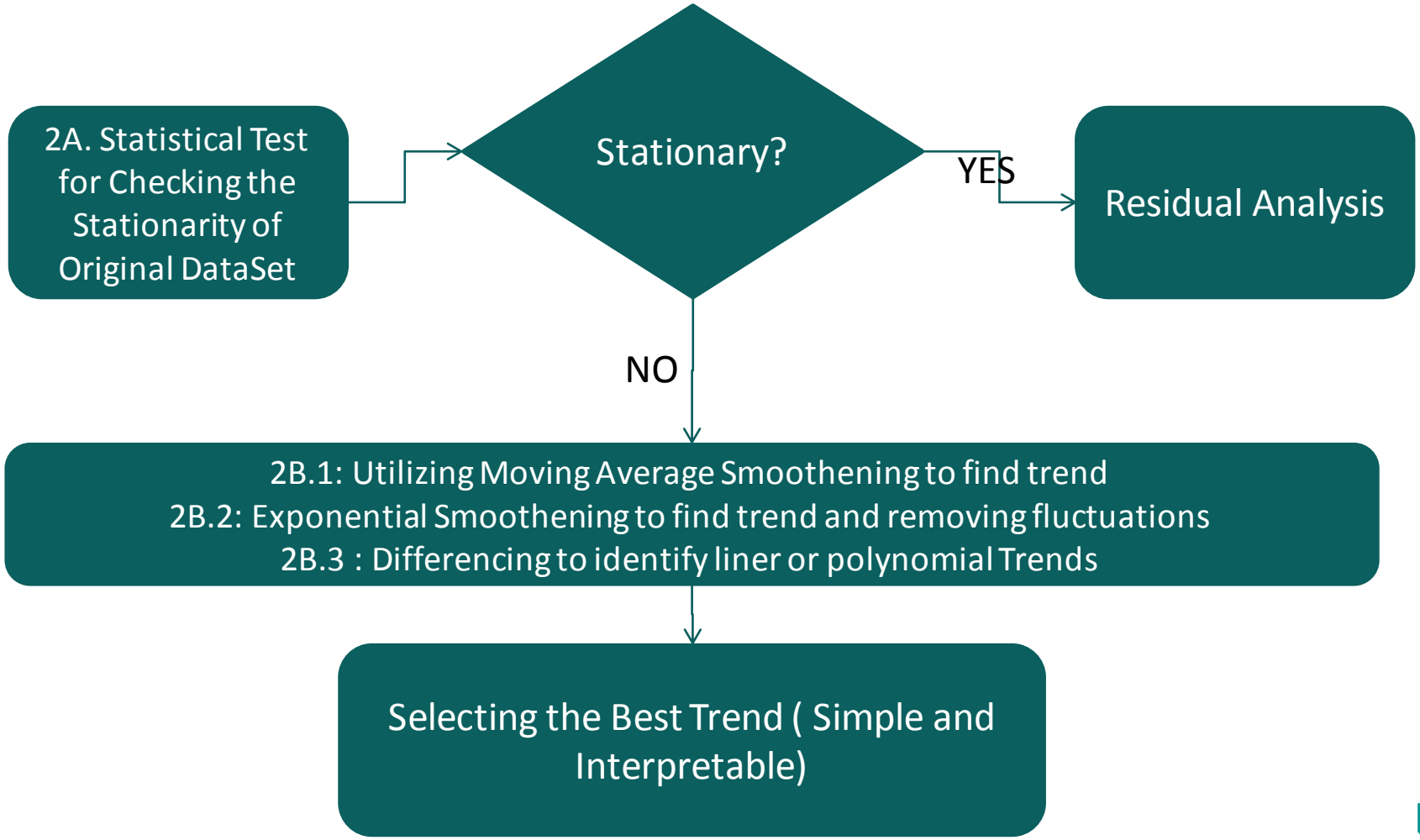
PREDICTING NUMBER OF FUTURE TRANSACTIONS

1.PREPROCESSING: Plotting, Optional Transformation, and Splitting for Supervised Learning



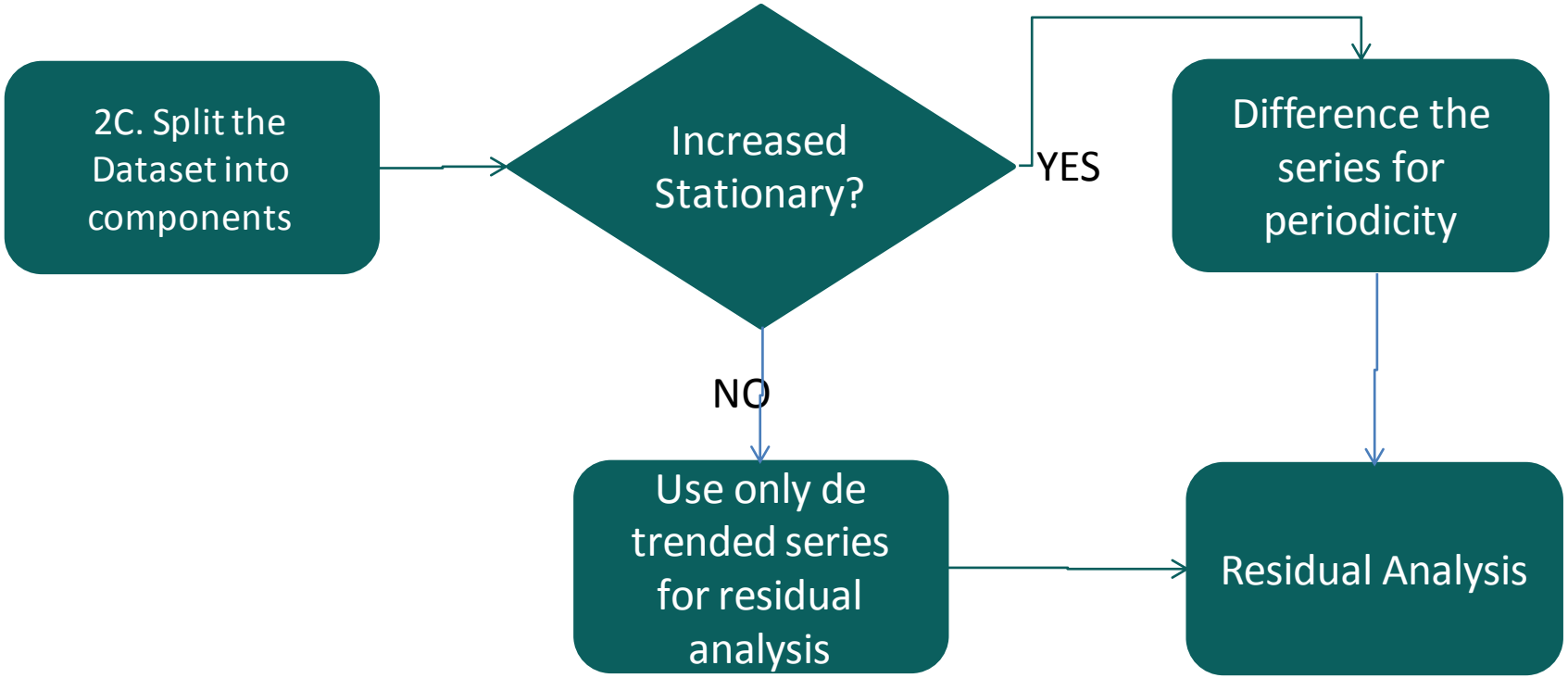
PREDICTING NUMBER OF FUTURE TRANSACTIONS

2. PROCESSING: Checking for Stationarity and Identifying the trends



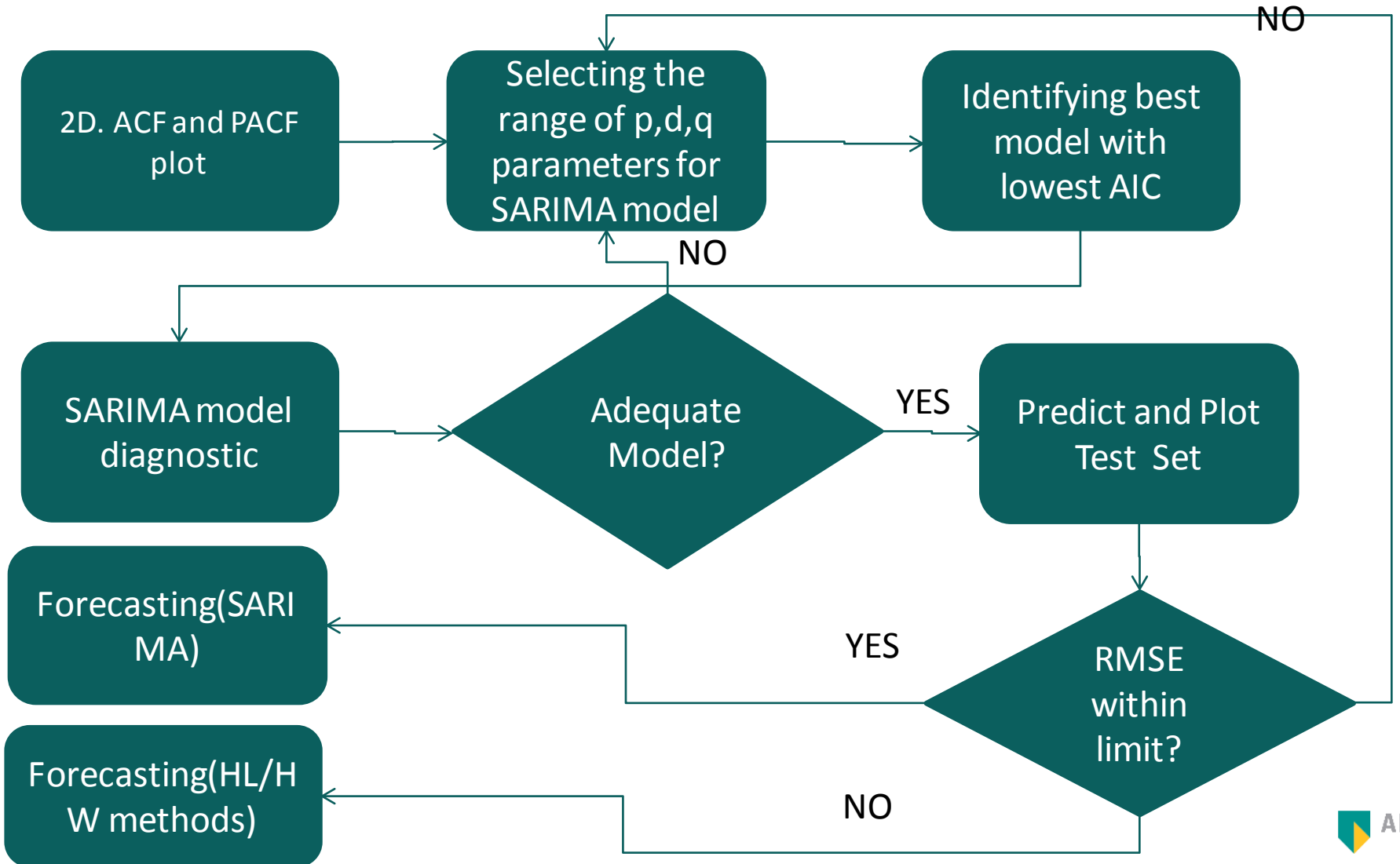
PREDICTING NUMBER OF FUTURE TRANSACTIONS

2. PROCESSING: Splitting the Dataset into Trend, Seasonal Component, and Residuals

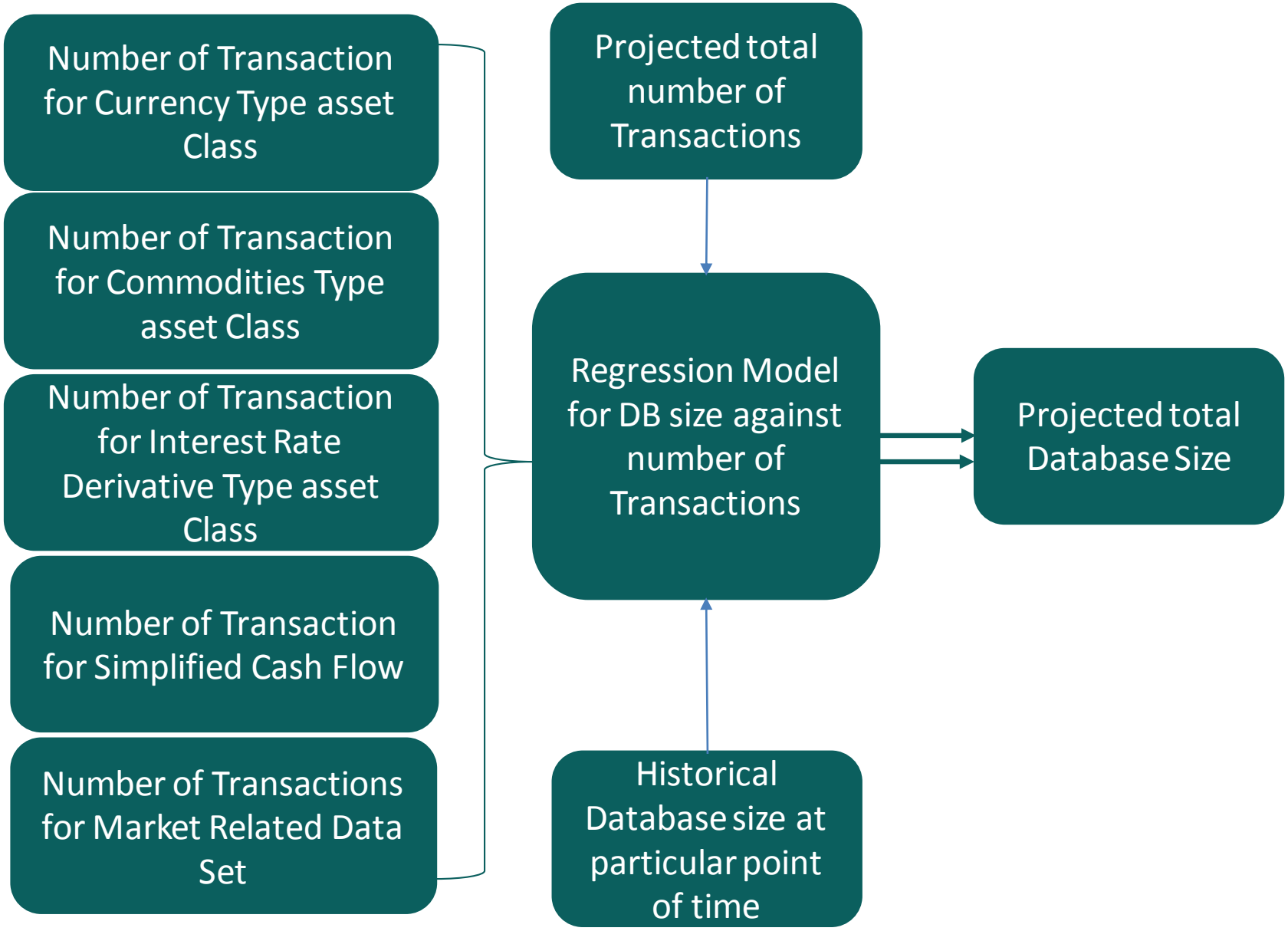


PREDICTING NUMBER OF FUTURE TRANSACTIONS

2:PROCESSING(cont.)&3FORECAASTING: Residual Analysis and Forecasting



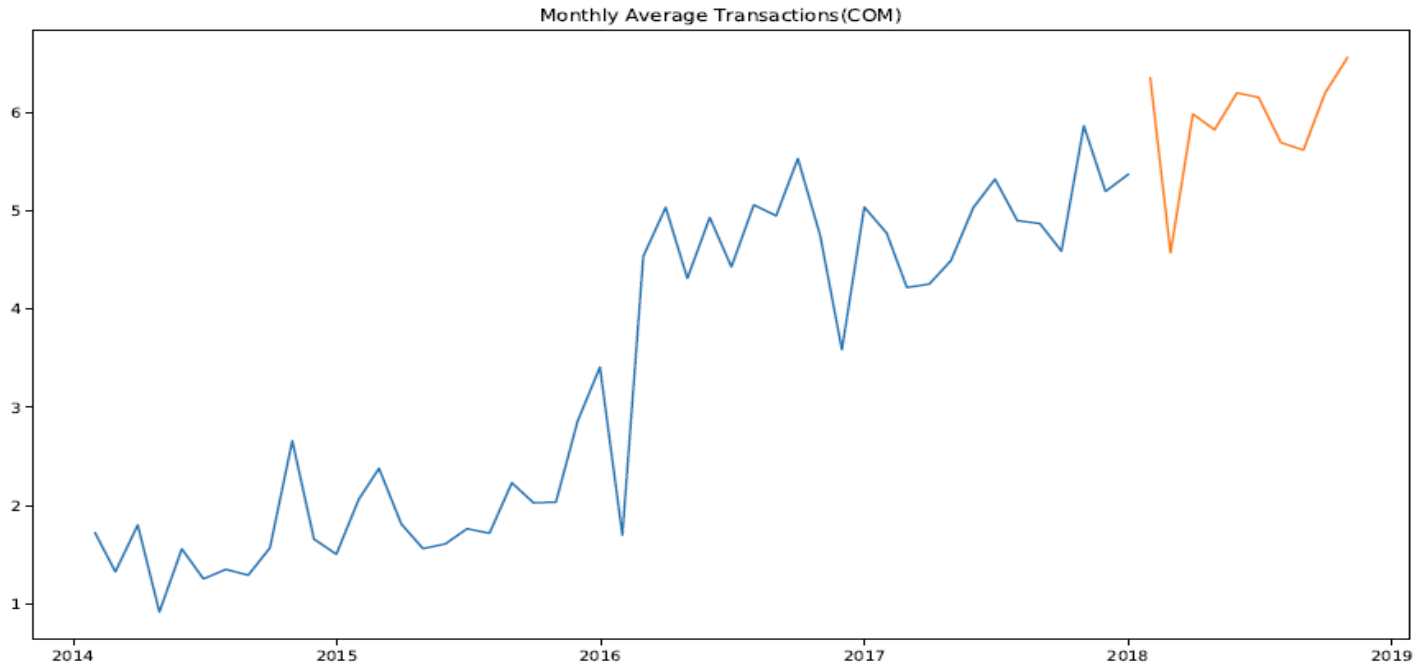
DB size for Number of Transactions



APPENDIX

Example: Forecasting number of Transactions for Commodity(COM) type Asset Class

1.Initial Plotting of Dataset (after converting to natural logarithmic scale)



Diagnostics

- Since the Variation is number of transactions is varying with time, we have applied natural logarithmic transformation to change the variability scale
- Existence of general upward trend

Example: Forecasting number of Transactions for Commodity(COM) type Asset Class

2.Checking for Stationarity of Existing Dataset

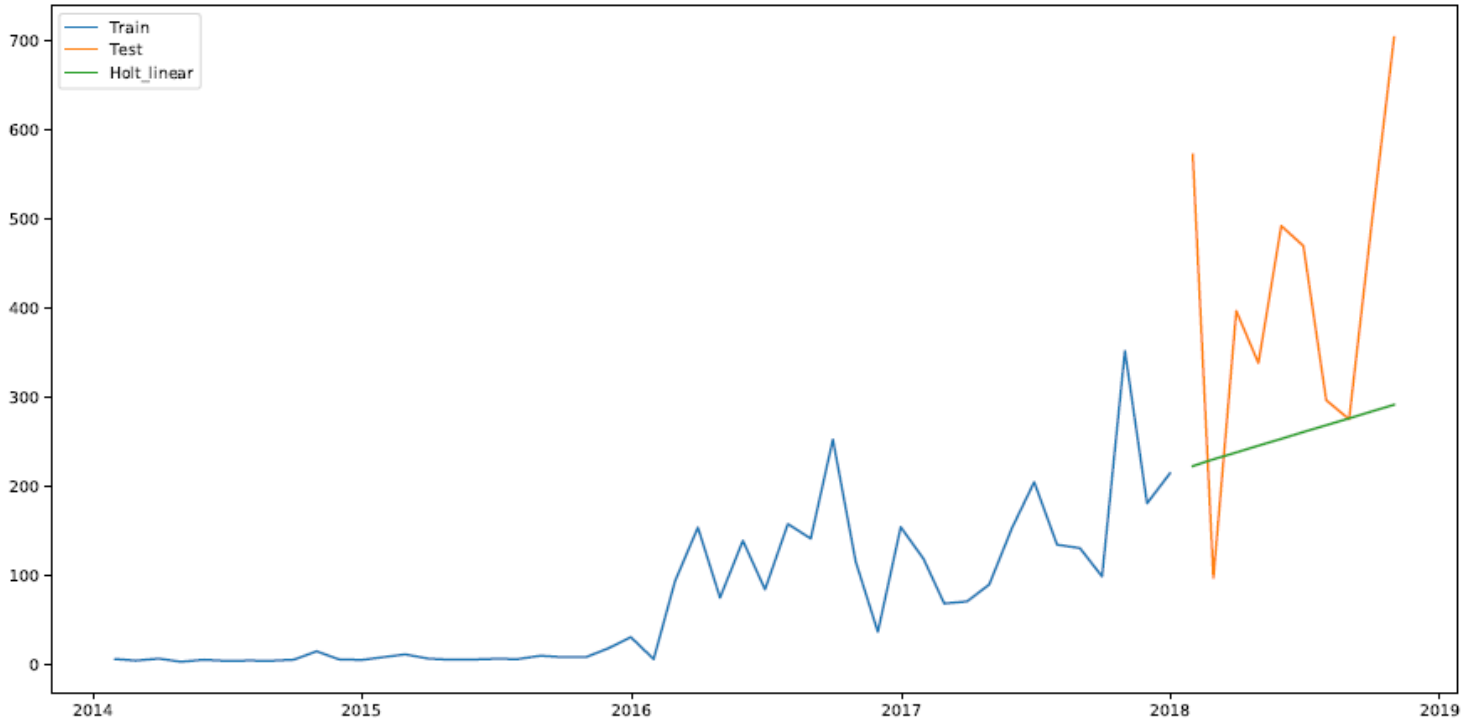
```
Results of Dickey-Fuller Test:
Test Statistic                  -0.425168
p-value                          0.905811
#Lags Used                       2.000000
Number of Observations Used     55.000000
Critical Value (1%)             -3.555273
Critical Value (5%)             -2.915731
Critical Value (10%)            -2.595670
dtype: float64
```

Diagnostics

- We cannot reject the Null Hypothesis of ADF test, hence concluded that Dataset is non stationary

Example: Forecasting number of Transactions for Commodity(COM) type Asset Class

3.Testing a Holt Linear Approximation for forecasting

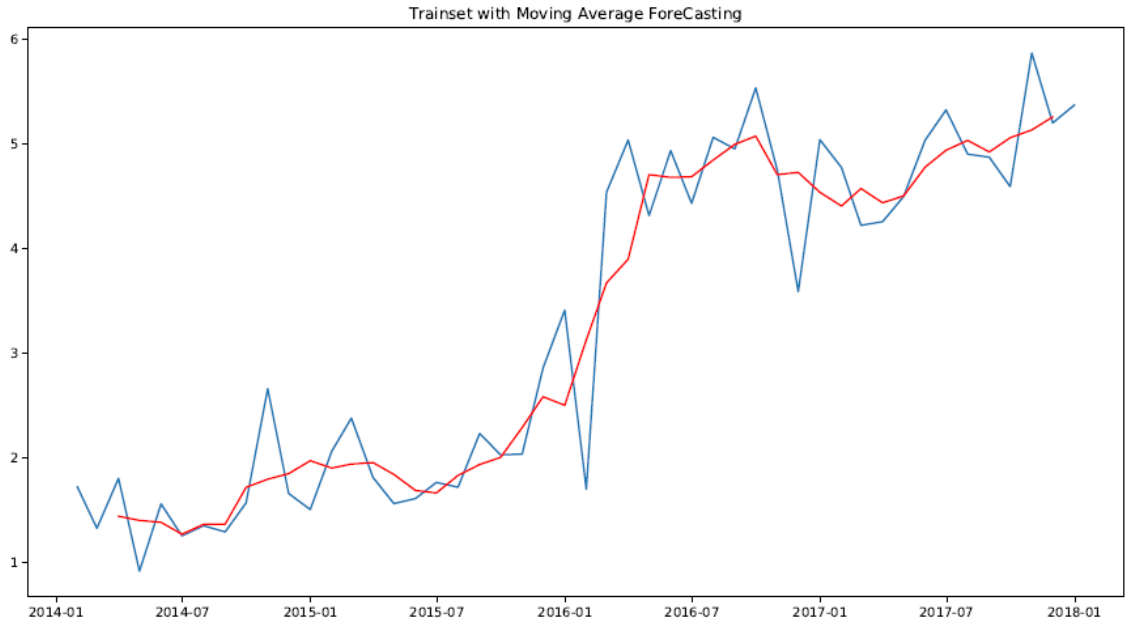


Diagnostics

- Holt linear model provides a reasonable estimate with Mean Squared Error of 221.34

Example: Forecasting number of Transactions for Commodity(COM) type Asset Class

4A. Checking for Trend (Moving average Method)



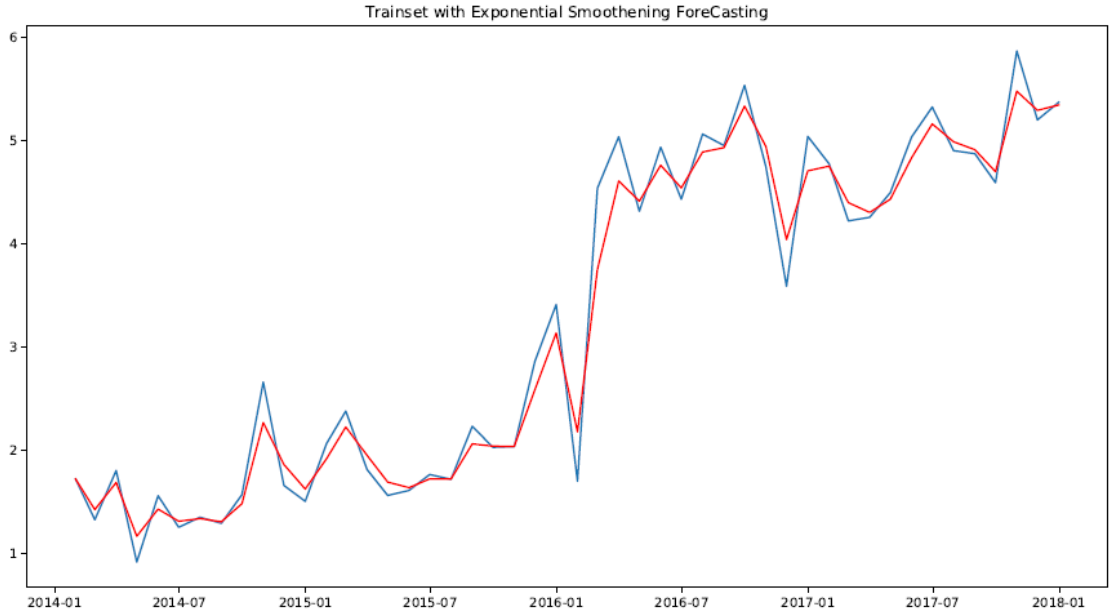

```
Results of Dickey-Fuller Test:
Test Statistic                -9.044735e+00
p-value                       5.022276e-15
#Lags Used                    1.000000e+00
Number of Observations Used   4.300000e+01
Critical Value (1%)           -3.592504e+00
Critical Value (5%)           -2.931550e+00
Critical Value (10%)          -2.604066e+00
dtype: float64
```

Diagnostics

- Presence of an upward trend
- Removal of Trend renders the Dataset Stationary (indicative of absence of no seasonal component)

Example: Forecasting number of Transactions for Commodity(COM) type Asset Class

4A. Checking for Trend (Exponential Smoothening)



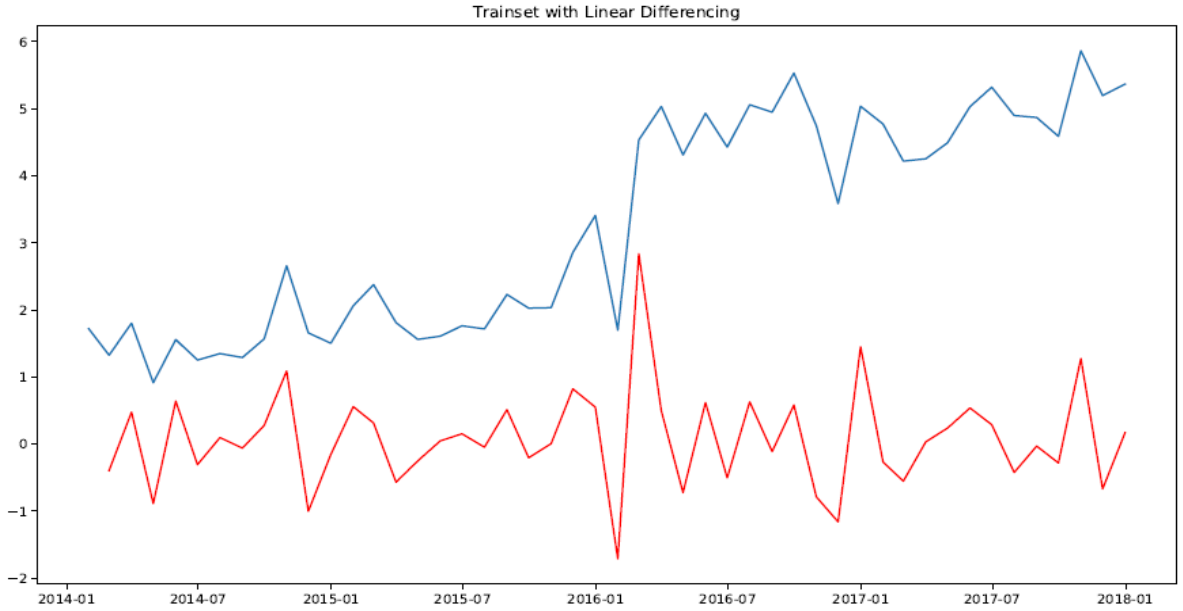

```
Results of Dickey-Fuller Test:
Test Statistic                  -6.750889e+00
p-value                         2.953126e-09
#Lags Used                      1.000000e+00
Number of Observations Used     4.600000e+01
Critical Value (1%)             -3.581258e+00
Critical Value (5%)             -2.926785e+00
Critical Value (10%)            -2.601541e+00
dtype: float64
```

Diagnostics

- Presence of an upward trend
- Removal of Trend renders the Dataset Stationary (indicative of absence of no seasonal component)

Example: Forecasting number of Transactions for Commodity(COM) type Asset Class

4A. Checking for Trend (Linear Trend)



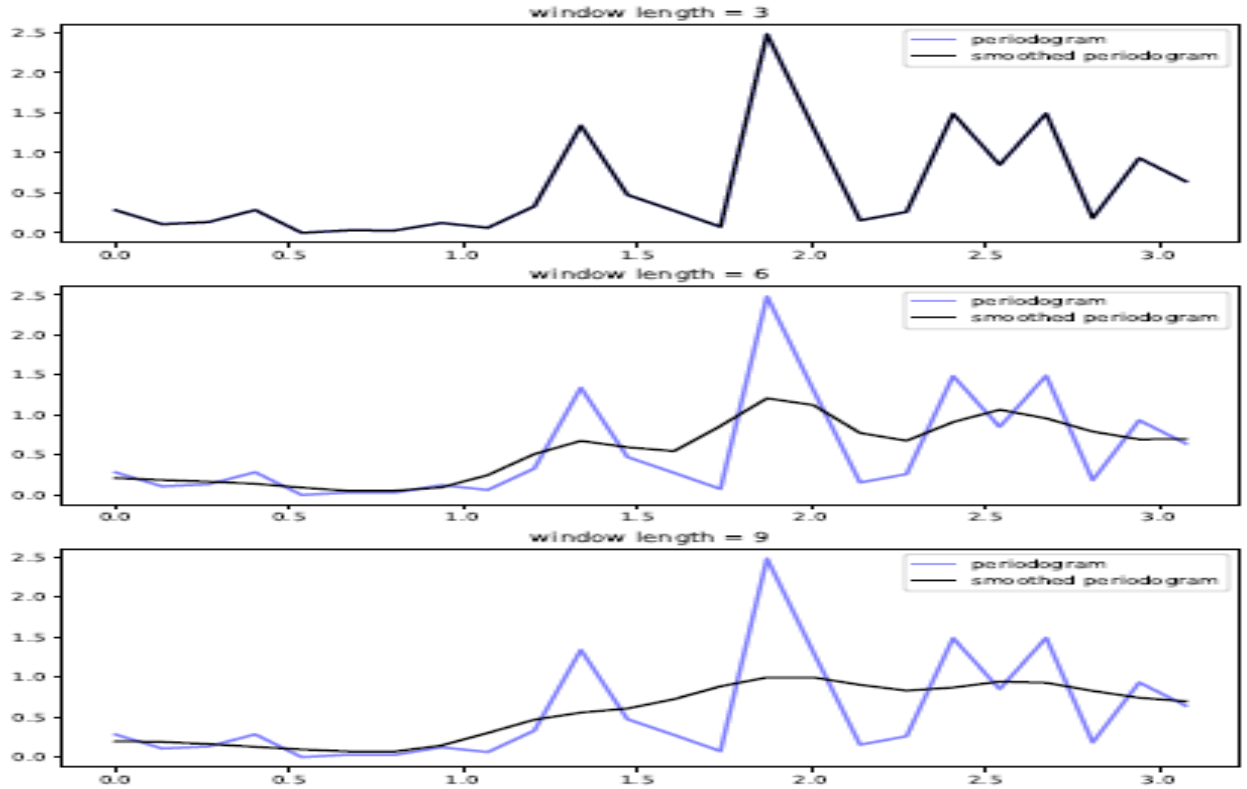

```
Results of Dickey-Fuller Test:
Test Statistic                -8.183560e+00
p-value                        8.010773e-13
#Lags Used                     1.000000e+00
Number of Observations Used    4.500000e+01
Critical Value (1%)           -3.584829e+00
Critical Value (5%)           -2.928299e+00
Critical Value (10%)          -2.602344e+00
dtype: float64
```

Diagnostics

- Presence of an upward trend
- Removal of Trend renders the Dataset Stationary (indicative of absence of no seasonal component)
- We can use the linear trend for modelling as this will reduce bias in predicted model compared to the previous two methods

Example: Forecasting number of Transactions for Commodity(COM) type Asset Class

5. Checking for Seasonality (Periodogram Analysis)



Diagnostics

- There is no specific Frequency that is largely responsible for the signature of the variation
- Smoothening reveals that there is no significant periodicity

Example: Forecasting number of Transactions for Commodity(COM) type Asset Class

5. Checking for Stationarity of Residuals (Results from TS decomposition)

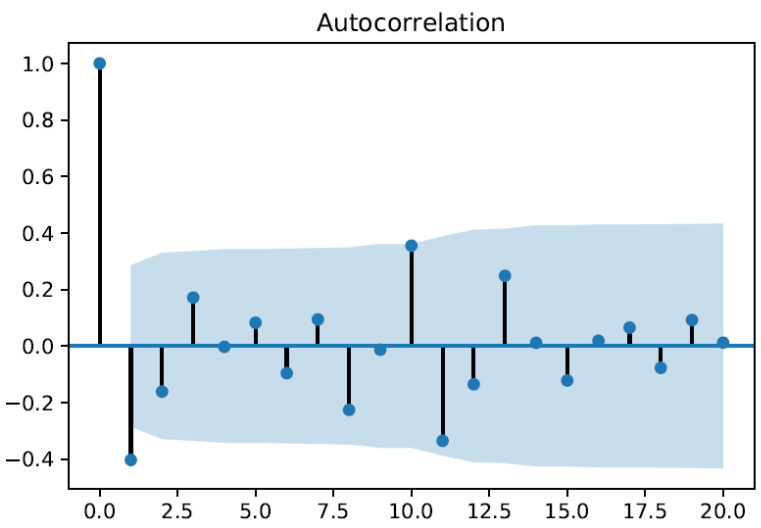
```
Results of Dickey-Fuller Test:
Test Statistic                  -3.191501
p-value                          0.020482
#Lags Used                       7.000000
Number of Observations Used     28.000000
Critical Value (1%)             -3.688926
Critical Value (5%)             -2.971989
Critical Value (10%)            -2.625296
dtype: float64
```

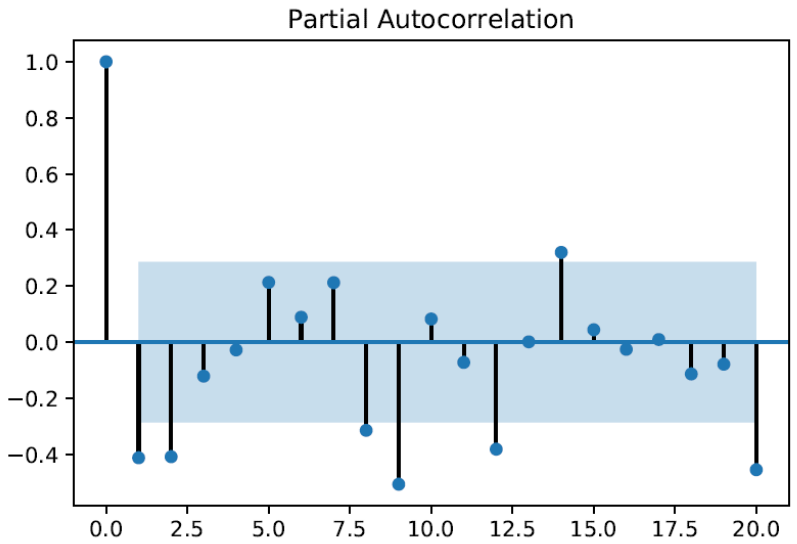
Diagnostics

- The results further strengthens the assumption that residuals are stationary

Example: Forecasting number of Transactions for Commodity(COM) type Asset Class

5A. Residual Modelling (AR and MA modelling)





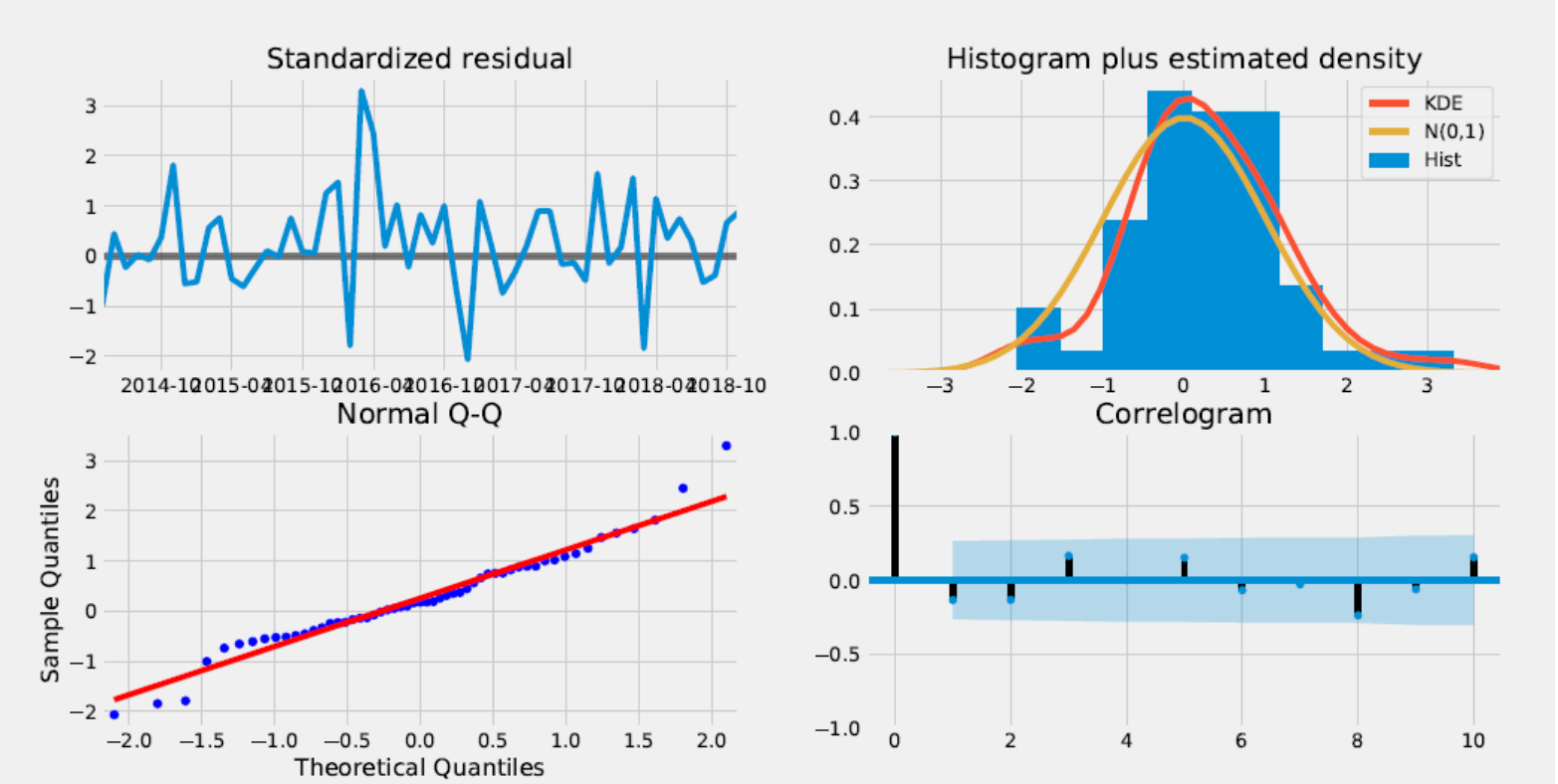
Diagnostics

- The plots suggests three possible models
  - ARMA(0,1,1) model
  - MA(1,0,1) model
  - ARMA(1,1,1) model
  - ARMA(0,1,1) model is selected based upon model diagonisitics


5.1_ARIMA Models Diagonistics.txt

Example: Forecasting number of Transactions for Commodity(COM) type Asset Class

5B. Comparison among three models



Diagnostics

- ARMA(0,1,1) model can be used for modelling the residuals as it provides the least AIC among all the models
- All models more or less approximates the residuals as Normal Distribution with 0 mean

Example: Forecasting number of Transactions for Commodity(COM) type Asset Class

6. Plotting the predicted Results for Test Set

Diagnostics

- The model fairly approximates the test result set.

Example: Forecasting number of Transactions for Commodity(COM) type Asset Class

6. Forecasted Values Set (with 95% confidence interval)

DateTime	Actual	LowerBound	UpperBound
2018-01-31	572.739130	96.366266	310.730319
2018-02-28	96.761905	89.422482	317.674103
2018-03-31	396.909091	82.877606	324.218979
2018-04-30	338.238095	76.669892	330.426693
2018-05-31	492.478261	70.752049	336.344536
2018-06-30	470.000000	65.086903	342.009682
2018-07-31	296.565217	59.644609	347.451976
2018-08-31	275.130435	54.400768	352.695817
2018-09-30	491.105263	49.335135	357.761450
2018-10-31	704.565217	44.430689	362.665896
2018-11-30	NaN	39.672957	367.423629
2018-12-31	NaN	35.049510	372.047075
2019-01-31	NaN	30.549583	376.547002
2019-02-28	NaN	26.163774	380.932811
2019-03-31	NaN	21.883818	385.212767
2019-04-30	NaN	17.702401	389.394184
2019-05-31	NaN	13.613016	393.483569
2019-06-30	NaN	9.609841	397.486744

****

The Root Mean Squared Error of Test Set(2018 Jan to 2018 Oct) is 197.63

Diagnostics

- All test Values are within 95% confidence interval range barring few extreme observations
- MSE for the model is 197.63 (better than Holt Linear Approximation model)

Current Status

- Time Series Modelling of SCF and IRD type of transactions
- Time Series Modelling of Market Dataset
- Linear Regression between Number of Transactions and DB size (for final volume prediction)

TASK(s)			Status
Time Series Modelling to forecast number of transactions	Asset Based Transactions	CURR (Currency)	Complete
		COM(Commodity)	Complete
		IRD(Interest Rate Derivatives)	In Progress
		SCF (Simplified CashFlow)	In Progress
	Market Data Set	Normal Data Set & Binary Data Set	Complete
Linear (Multiple) Regression modelling to check for the DB size for individual types of transactions	Asset Based Transactions & Market Related Transactions		Waiting for input (Historical DB size)

Infosys®
Navigate your next