BANA7050 Forecasting and Time Series Methods

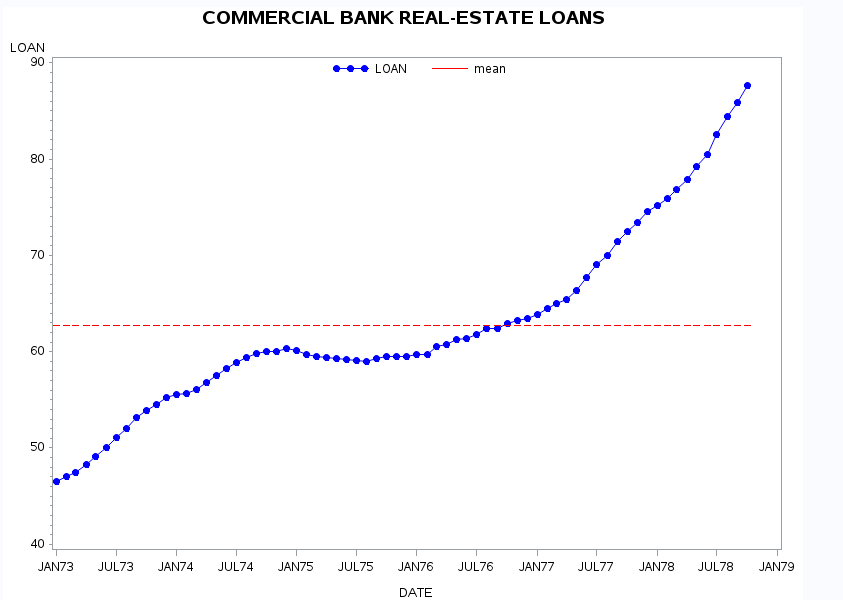
Case Study #2

By: Siva Ramakrishnan M#: M10669365

# Model Identification

First step in the model identification would be plotting the Loans vs the month.

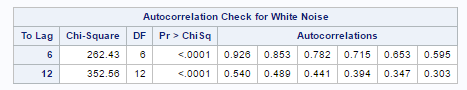
Looking at the plot the data there is definitely non-stationarity observable in the data



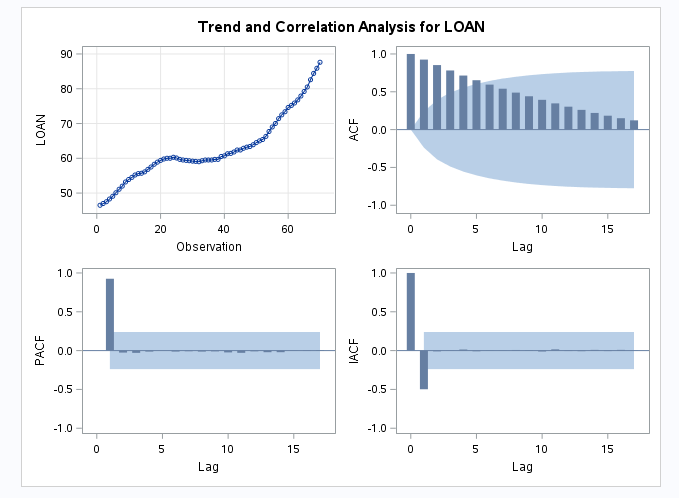
Next step would be checking if the data follows white noise using the Portmanteau Q test.

The test checks the Null hypothesis that the data is white noise.

P-values outputted by SAS prove that we can reject the null hypothesis



Next, we Look at the ACF and PACF values for upto 17 lags (as there are 70 obs we would only need to examine around 70/4 lags)



The following inferences can be drawn from the above plot

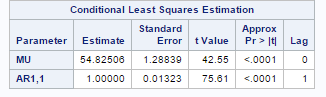
1. The ACF values seem to be decreasing very slowly suggesting that the data is non-stationary.

2. There is only one significant PACF Value at lag of 1.

**Building Initial AR(1) model:**

The decreasing ACF and the PACF getting cut off at lag 1 suggest that AR(1) may to be a good model for the data.

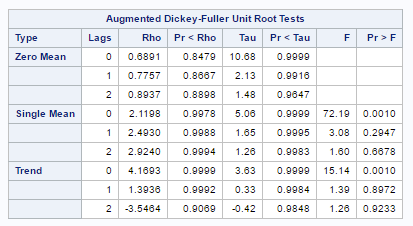
On estimating the parameters by considering AR(1) model we get the parameter estimate to be equal to unit root a very strong indication that the model is non-stationary.



**Non-Stationarity Test:**

Augmented Dickey-Fuller test is used for checking Nonstationary data.

The null hypothesis of the test is that the data is Homogenous Non-Stationarity.



We are getting very high P-values. Hence we cannot reject the null hypothesis that data is Homogenous Non-Stationary.

Based on the ADF test, Parameter estimate of AR(1) and the plot we can conclude that the data is non-stationary.

**Code for Performing ADF and estimating AR(1) parameters :**

PROC ARIMA;

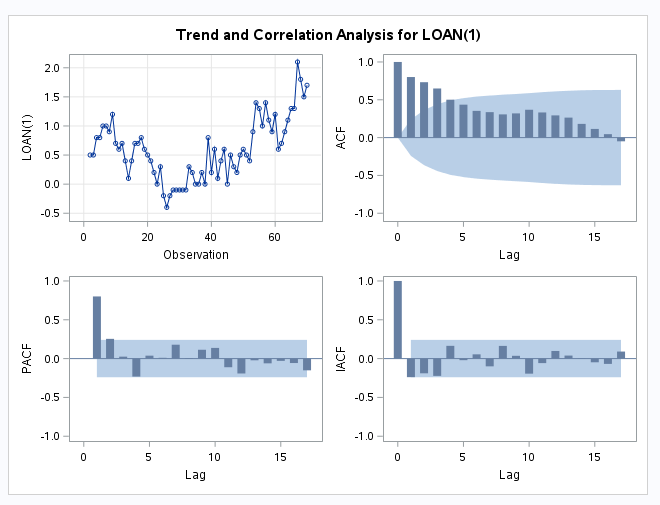
IDENTIFY VAR=SAVING STATIONARITY=(ADF);

ESTIMATE P=1;

RUN;

**First Order difference**

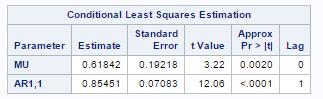
Since the data was identified as being non-stationary the next step would be looking at the first order difference values



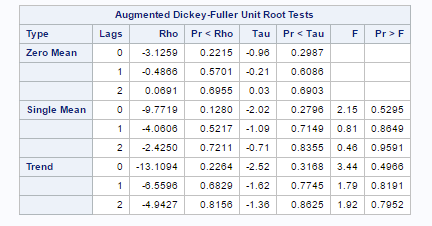
The Plot for the first order difference again seems to be non-stationary.

The ACF also seems to be decreasing very slowly suggesting data is non-stationary

Based on decreasing ACF if we fit a AR(1) model on the first order differences we get the below parameter estimates



A high parameter estimate close to 1 again suggests that the data is non-stationary.



ADF test suggest that the first order difference is non-stationary since all P-values are greater than 0.05

Based on the ADF test, Parameter estimate of AR(1) and the plot we can conclude that the first order difference is again non-stationary.

**Code for Performing ADF and estimating AR(1) parameters on First Order Differences:**

PROC ARIMA;

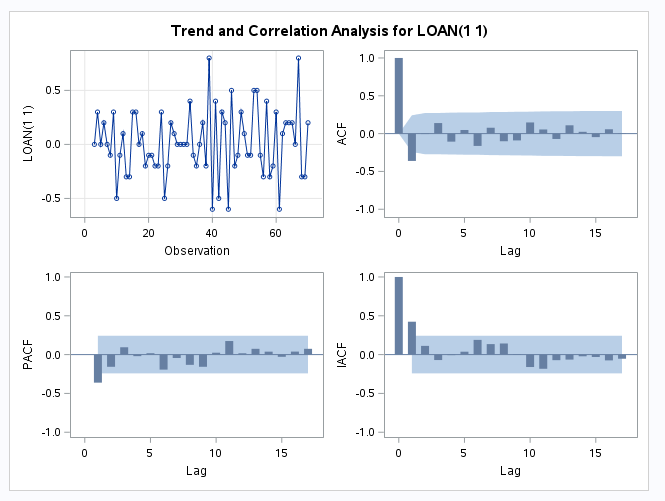
IDENTIFY VAR=LOAN(1) STATIONARITY=(ADF);

ESTIMATE P=1;

RUN;

**Second Order difference**

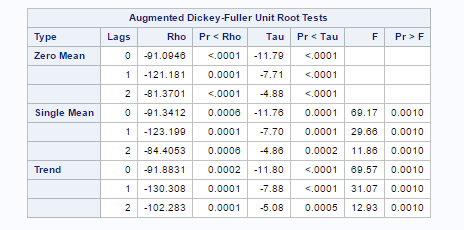
Since the first order difference was identified as being non-stationary the next step would be looking at the second order difference values



Looking at the ACF and PACF plots, ACF seems to have a single spike at lag(1) and PACF seems to be exponentially decreasing suggesting a MA(1) model for second order difference.

Data appears to be non-stationary looking at the plots.

In the ADF test, we see that P-values are less than 0.05 so we can reject the null hypothesis that the data is non-stationary



Hence we can conclude that the second order difference is Non-stationary and MA(1) model seems to be good model on the second order difference. We can finalize on the final model on the second order differences based on MINIC.

**Code for Performing ADF and estimating AR(1) parameters on Second Order Differences:**

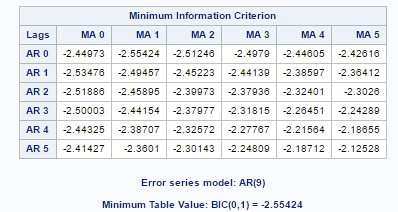
PROC ARIMA;

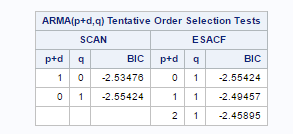
IDENTIFY VAR=LOAN(1,1) STATIONARITY=(ADF);

RUN;

**MINIC, ESACF and SCAN :**

Looking at the MINIC table for the Second order differentials, we see that MA(1) model gives us the least value of BIC





SCAN and ESACF also seem to suggest MA(1) model

**Code for Getting MINIC, ESACF and SCAN values :**

PROC ARIMA;

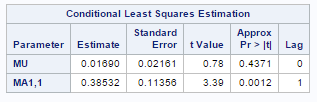
IDENTIFY VAR=LOANS(2,1) STATIONARITY=(ADF) MINIC ESACF SCAN;

RUN;

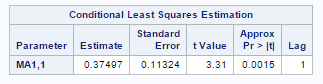
Thus considering all the previous factors including the ACF/PACF plots, MINIC,ESACF and SCAN, a ARIMA(0,2,1) models seems to fit the original data.

# Estimation of model parameters

The estimated mean of the selected ARIMA(0,2,1) model is 0.01690 and phi1 value is .38532 using the conditional least squares method. P-Value of mean suggests that it can be assumed to be zero.



**Final Model** (using noconstant option to make mean zero)



**Code for estimating model parameters and doing residual diagnostics:**

PROC ARIMA;

IDENTIFY VAR=LOAN(1,1);

ESTIMATE q=1 NOCONSTANT;

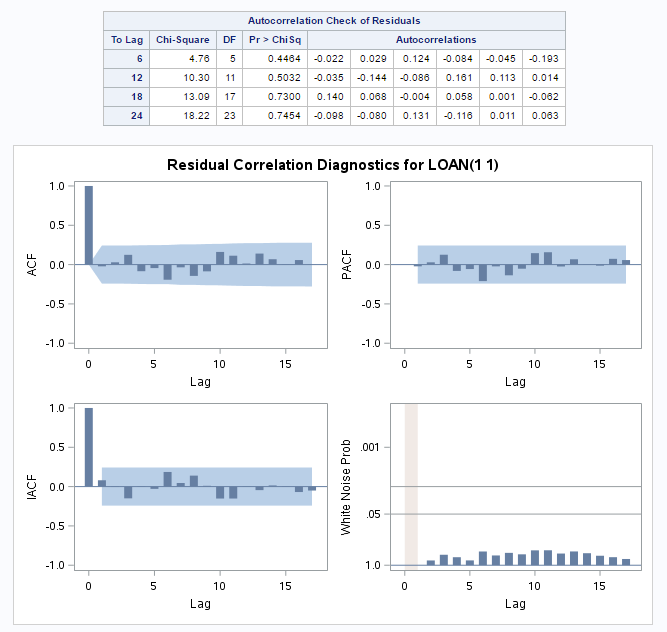
RUN;

**Diagnostics:**

The correlation diagnostics agree that the residual is white noise since there are no spikes in the plots.

The P-values in the table are greater than 0.05 suggesting that we cannot reject the null hypothesis that correlation is zero.

Thus the ARIMA(0,2,1) model is validated.

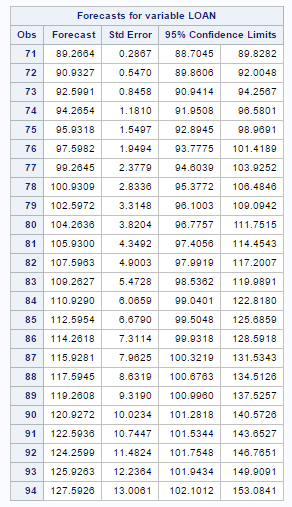


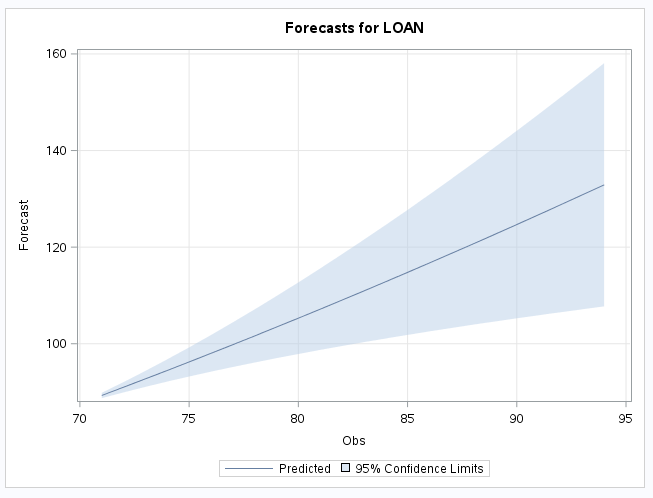
# Forecasting

For forecasting the values for the next 2 years we need to forecast 24 observations which correspond to a month. The forecasted values are given below.

We observe that when we are predicting the value in the immediate future the confidence band is small i.e; for observation number 71 we see a small confidence level of (88.73, 89.8571).

As we predict further into the future the standard error and the range of confidence limits increases.





**Code for forecasting the values:**

PROC ARIMA DATA=CASE;

IDENTIFY VAR=LOAN(1,1);

ESTIMATE Q=1 NOCONSTANT;

FORECAST LEAD=24;

RUN;