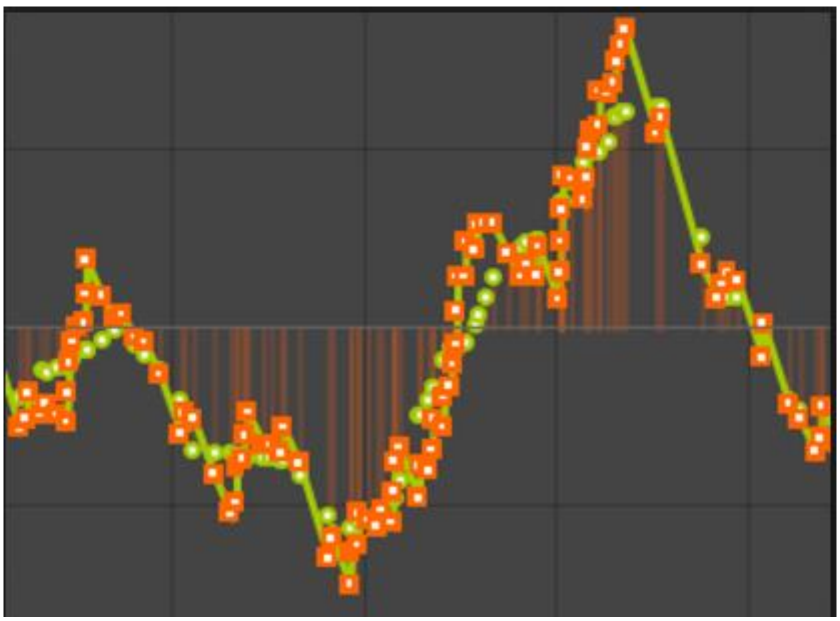

## ABSTRACT

White noise & Stationarity check, data transformation, model identification, model selection, residual diagnosis , parameter estimation, and forecasting is done on real-estate Loans data from commercial banks.

Author: Swapnil Sharma
M10743959

# Table of Contents

## Data Set:

The monthly volume of commercial bank real-estate loans, in billions of dollars, from January 1973 to October 1978, a total of 70 observations. The data is derived from the reports to the Federal Reserve System from large commercial banks.

**Code:**

```
DATA CASE;
   DO YEAR=73 TO 78;
     DO MONTH=1 TO 12;
       DATE=MDY(MONTH,1,YEAR);
                     one=1;
       INPUT LOAN @@;
                     Output;
     END;
   END;
KEEP DATE LOAN one;
FORMAT DATE MONYY5.;
Title 'COMMERCIAL BANK REAL-ESTATE LOANS';
CARDS;
46.5 47 47.5 48.3 49.1 50.1 51.1 52 53.2 53.9 54.5 55.2 55.6 55.7 56.1
56.8 57.5 58.3 58.9 59.4 59.8 60 60 60.3 60.1 59.7 59.5 59.4 59.3 59.2
59.1 59 59.3 59.5 59.5 59.5 59.7 59.7 60.5 60.7 61.3 61.4 61.8
62.4 62.4 62.9 63.2 63.4 63.9 64.5 65 65.4 66.3 67.7 69 70 71.4
72.5 73.4 74.6 75.2 75.9 76.8 77.9 79.2 80.5 82.6 84.4 85.9 87.6
;
PROC PRINT;
RUN;
```

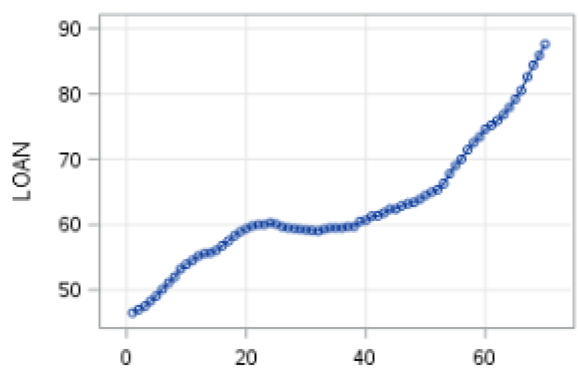

*Figure 1 Loan across time*

We see increasing loan amount given across the years as per above figure.

## Chapter1: White Noise check for original data

White Noise: We desire our original data to not follow white noise characteristics so that time series analysis can be done.

**Code:**

```
/* Null Hypothesis of white noise for original data testing*/
proc arima data=case;
identify var=LOAN;
run;
```

Null Hypothesis: Time series is White Noise

Result: We reject the null hypothesis as P-value is insignificant which can be seen in the figure below

| Autocorrelation Check for White Noise | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 262.43 | 6 | <.0001 | 0.926 | 0.853 | 0.782 | 0.715 | 0.653 | 0.595 |
| 12 | 352.56 | 12 | <.0001 | 0.540 | 0.489 | 0.441 | 0.394 | 0.347 | 0.303 |

*Figure 2 Check for White Noise: Original data*

## Chapter2: Stationarity check for original data

Stationarity is required condition for time series analysis. Either original data should be stationary or it should be transformed by various methods and then analyze. Stationarity conditions imply constant mean, constant and constant variance. We can check for stationarity by looking at the ACF plot as well as do ADF test. IF the ACF is slowly decreasing then time series can be called non-stationary.

**Code:**

```
/* ADF test for stationarity */
proc arima data=case;
identify VAR=LOAN STATIONARITY=(ADF);
RUN;
```

ADF-Test

Null Hypothesis: Time series data is homogenous non-stationary (have unit root)

Result: We cannot reject the null hypothesis for ADF test as we have significant P-value. This indicates that our data is homogeneous non-stationary and we need to transform it to make stationary.

| Augmented Dickey-Fuller Unit Root Tests | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type | Lags | Rho | Pr < Rho | Tau | Pr < Tau | F | Pr > F |
| Zero Mean | 0 | 0.6891 | 0.8479 | 10.68 | 0.9999 | | |
| | 1 | 0.7757 | 0.8667 | 2.13 | 0.9916 | | |
| | 2 | 0.8937 | 0.8898 | 1.48 | 0.9647 | | |
| Single Mean | 0 | 2.1198 | 0.9978 | 5.06 | 0.9999 | 72.19 | 0.0010 |
| | 1 | 2.4930 | 0.9988 | 1.65 | 0.9995 | 3.08 | 0.2947 |
| | 2 | 2.9240 | 0.9994 | 1.26 | 0.9983 | 1.60 | 0.6678 |
| Trend | 0 | 4.1693 | 0.9999 | 3.63 | 0.9999 | 15.14 | 0.0010 |
| | 1 | 1.3936 | 0.9992 | 0.33 | 0.9984 | 1.39 | 0.8972 |
| | 2 | -3.5464 | 0.9069 | -0.42 | 0.9848 | 1.26 | 0.9233 |

*Figure 3 ADF Test on original data*

As we do not have zero or constant mean but increasing trend in the data we consider highlighted P-value to take decision

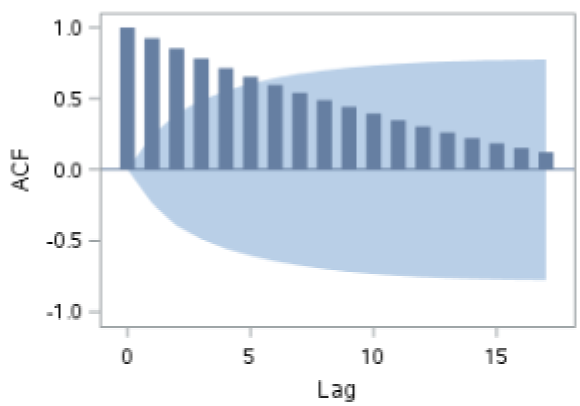


*Figure 4 ACF vs Lag Plot for original data*

## Chapter 3: Transformation and stationarity check

Now we take first order difference of the time series and again do the ADF test to check stationarity of transformed data.

**Code:**

```
/*First order Transformation and stationarity test*/
proc arima data=case;
identify VAR=LOAN(1) STATIONARITY=(ADF);
RUN;
```

| Augmented Dickey-Fuller Unit Root Tests | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type | Lags | Rho | Pr < Rho | Tau | Pr < Tau | F | Pr > F |
| Zero Mean | 0 | -3.1259 | 0.2215 | -0.96 | 0.2987 | | |
| | 1 | -0.4866 | 0.5701 | -0.21 | 0.6086 | | |
| | 2 | 0.0691 | 0.6955 | 0.03 | 0.6903 | | |
| Single Mean | 0 | -9.7719 | 0.1280 | -2.02 | 0.2796 | 2.15 | 0.5295 |
| | 1 | -4.0606 | 0.5217 | -1.09 | 0.7149 | 0.81 | 0.8649 |
| | 2 | -2.4250 | 0.7211 | -0.71 | 0.8355 | 0.46 | 0.9591 |
| Trend | 0 | -13.1094 | 0.2264 | -2.52 | 0.3168 | 3.44 | 0.4966 |
| | 1 | -6.5596 | 0.6829 | -1.62 | 0.7745 | 1.79 | 0.8191 |
| | 2 | -4.9427 | 0.8156 | -1.36 | 0.8625 | 1.92 | 0.7952 |

*Figure 5 ADF Test Output for First Order Difference*

I have used only ADF test to make my decision about stationarity. As all the P-values are significant we can conclude the data is not stationary. This can be also seen from ACF plot as below.

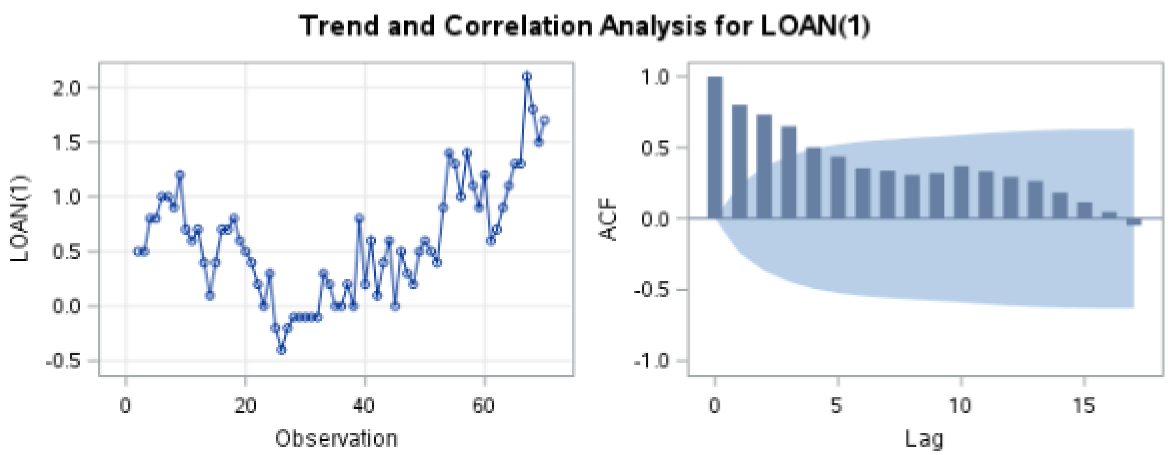


*Figure 6 First Order Difference: Time series and ACF vs Lag plot*

## Chapter 4: Second order Transformation and ADF test

As we didn't get stationary time series data after first order transformation we take second order transformation and check. We might try log transformation and square root transformation if we do not get stationary data after this Chapter.

**Code:**

```
/*second order Transformation and stationarity test*/
proc arima data=case;
identify VAR=LOAN(1,1) STATIONARITY=(ADF);
RUN;
```

| Augmented Dickey-Fuller Unit Root Tests | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type | Lags | Rho | Pr < Rho | Tau | Pr < Tau | F | Pr > F |
| Zero Mean | 0 | -91.0946 | <.0001 | -11.79 | <.0001 | | |
| | 1 | -121.181 | 0.0001 | -7.71 | <.0001 | | |
| | 2 | -81.3701 | <.0001 | -4.88 | <.0001 | | |
| Single Mean | 0 | -91.3412 | 0.0006 | -11.76 | 0.0001 | 69.17 | 0.0010 |
| | 1 | -123.199 | 0.0001 | -7.70 | 0.0001 | 29.66 | 0.0010 |
| | 2 | -84.4053 | 0.0006 | -4.86 | 0.0002 | 11.86 | 0.0010 |
| Trend | 0 | -91.8831 | 0.0002 | -11.80 | <.0001 | 69.57 | 0.0010 |
| | 1 | -130.308 | 0.0001 | -7.88 | <.0001 | 31.07 | 0.0010 |
| | 2 | -102.283 | 0.0001 | -5.08 | 0.0005 | 12.93 | 0.0010 |

*Figure 7 ADF Test Output for Second Order Difference*

As the P-values are insignificant we reject the null hypothesis of homogeneous non-stationarity and conclude the second order difference is stationary for the time series. The plot of second order time-series data and ACF vs Lag data can be seen below.

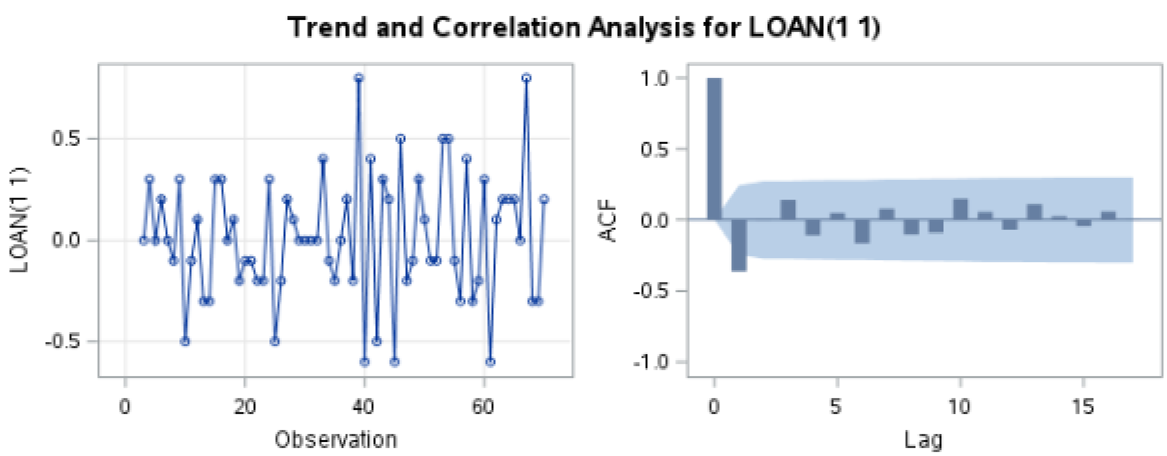


*Figure 8 Second Order Difference: Time series data and ACF vs Lag plot*

## Chapter 5: Model Identification

We have finally transformed original non-stationary data to stationary data and now we need to determine model. We use MINIC, ESACF and SCAN outputs to determine our model.

**Code:**

```
/*Model identification*/
proc arima data=case;
identify VAR=LOAN(1,1) MINIC ESACF SCAN;
RUN;
```

Minimum Table Value: BIC(0,1) = -2.55424

| ARMA(p+d,q) Tentative Order Selection Tests | | | | | |
|---|---|---|---|---|---|
| SCAN | | | ESACF | | |
| p+d | q | BIC | p+d | q | BIC |
| 1 | 0 | -2.53476 | 0 | 1 | -2.55424 |
| 0 | 1 | -2.55424 | 1 | 1 | -2.49457 |
| | | | 2 | 1 | -2.45895 |

(5% Significance Level)

*Figure 9 MINIC Output*

Here least BIC value suggests the model to be MA(1). This can also be checked from the outputs of SCAN and ESACF. The highlighted cell shows the model to be selected in the figure below.

Squared Canonical Correlation Estimates

| Lags | MA 0 | MA 1 | MA 2 | MA 3 | MA 4 | MA 5 |
|---|---|---|---|---|---|---|
| AR 0 | 0.1313 | <.0001 | 0.0204 | 0.0133 | 0.0027 | 0.0314 |
| AR 1 | 0.0239 | 0.0218 | 0.0187 | 0.0024 | 0.0132 | 0.0319 |
| AR 2 | 0.0098 | 0.0008 | <.0001 | 0.0169 | 0.0062 | 0.0038 |
| AR 3 | 0.0006 | <.0001 | 0.0024 | 0.0164 | 0.0001 | 0.0101 |
| AR 4 | <.0001 | 0.0005 | 0.0133 | 0.0106 | 0.0148 | 0.0008 |
| AR 5 | 0.0461 | 0.0397 | 0.0229 | 0.0215 | 0.0121 | 0.0075 |

SCAN Chi-Square[1] Probability Values

| Lags | MA 0 | MA 1 | MA 2 | MA 3 | MA 4 | MA 5 |
|---|---|---|---|---|---|---|
| AR 0 | 0.0020 | 0.9644 | 0.3003 | 0.4148 | 0.7189 | 0.2185 |
| AR 1 | 0.2033 | 0.2564 | 0.2859 | 0.7125 | 0.4074 | 0.2632 |
| AR 2 | 0.4206 | 0.8403 | 0.9497 | 0.3299 | 0.6209 | 0.7082 |
| AR 3 | 0.8377 | 0.9868 | 0.7351 | 0.3765 | 0.9510 | 0.5333 |
| AR 4 | 0.9470 | 0.8650 | 0.3715 | 0.5186 | 0.4458 | 0.8720 |
| AR 5 | 0.0846 | 0.1953 | 0.3035 | 0.3673 | 0.5361 | 0.6015 |

Extended Sample Autocorrelation Function

| Lags | MA 0 | MA 1 | MA 2 | MA 3 | MA 4 | MA 5 |
|---|---|---|---|---|---|---|
| AR 0 | -0.3614 | -0.0080 | 0.1394 | -0.1084 | 0.0477 | -0.1628 |
| AR 1 | -0.3700 | -0.0038 | 0.1242 | -0.0663 | -0.0259 | -0.1625 |
| AR 2 | 0.4573 | 0.0783 | 0.0133 | -0.0451 | -0.0399 | -0.0684 |
| AR 3 | 0.2525 | -0.0157 | 0.0304 | -0.1619 | 0.0069 | -0.0553 |
| AR 4 | 0.2874 | 0.0052 | 0.0947 | -0.1789 | 0.0398 | -0.0858 |
| AR 5 | 0.0339 | -0.0636 | 0.3565 | -0.3433 | -0.1484 | -0.0577 |

ESACF Probability Values

| Lags | MA 0 | MA 1 | MA 2 | MA 3 | MA 4 | MA 5 |
|---|---|---|---|---|---|---|
| AR 0 | 0.0029 | 0.9648 | 0.3062 | 0.4416 | 0.7325 | 0.2438 |
| AR 1 | 0.0025 | 0.9774 | 0.3392 | 0.5978 | 0.8389 | 0.2003 |
| AR 2 | 0.0002 | 0.5777 | 0.9239 | 0.7388 | 0.7959 | 0.6519 |
| AR 3 | 0.0418 | 0.9047 | 0.8318 | 0.3506 | 0.9685 | 0.7172 |
| AR 4 | 0.0215 | 0.9686 | 0.5152 | 0.3088 | 0.8235 | 0.5757 |
| AR 5 | 0.7877 | 0.6150 | 0.0097 | 0.0094 | 0.2927 | 0.6998 |

*Figure 10 SCAN & ESACF Output*

The highlighted cell suggests that MA (1) model should be tried for second order difference.

## Chapter 6: Parameter Estimates and Residual Diagnosis

We consider our model to follow MA(1) and estimate the parameter. Residual diagnosis is also done and we want our residuals to follow white noise and they should be normally distributed for our model to be valid.

We find the estimate of mean to be zero as p value is significant for the hypothesis that mu is zero and hence remove same from the model and rerun the code.

| Conditional Least Squares Estimation | | | | | |
|---|---|---|---|---|---|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > \|t\| | Lag |
| MU | 0.01690 | 0.02161 | 0.78 | 0.4371 | 0 |
| MA1,1 | 0.38532 | 0.11356 | 3.39 | 0.0012 | 1 |

*Figure 11 Mean Estimate is zero*

**Code:**

```
/* Without mu parameter estimate*/
proc arima data=case;
identify VAR=LOAN(1,1);
estimate q=1 noconstant;
RUN;
```

## Residual Diagnosis:

The autocorrelation check is done for residuals and the null hypothesis is that the residuals are white noise. We get significant p-value and hence cannot reject the null hypothesis of white noise. Moreover, they are normally distributed.

| Autocorrelation Check of Residuals | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 4.70 | 5 | 0.4541 | -0.024 | 0.033 | 0.131 | -0.081 | -0.039 | -0.188 |
| 12 | 10.30 | 11 | 0.5037 | -0.031 | -0.141 | -0.083 | 0.165 | 0.116 | 0.019 |
| 18 | 13.29 | 17 | 0.7164 | 0.146 | 0.072 | 0.001 | 0.063 | 0.005 | -0.058 |
| 24 | 18.45 | 23 | 0.7327 | -0.092 | -0.074 | 0.138 | -0.111 | 0.018 | 0.072 |

*Figure 12 White Noise Hypothesis Testing of Residuals*

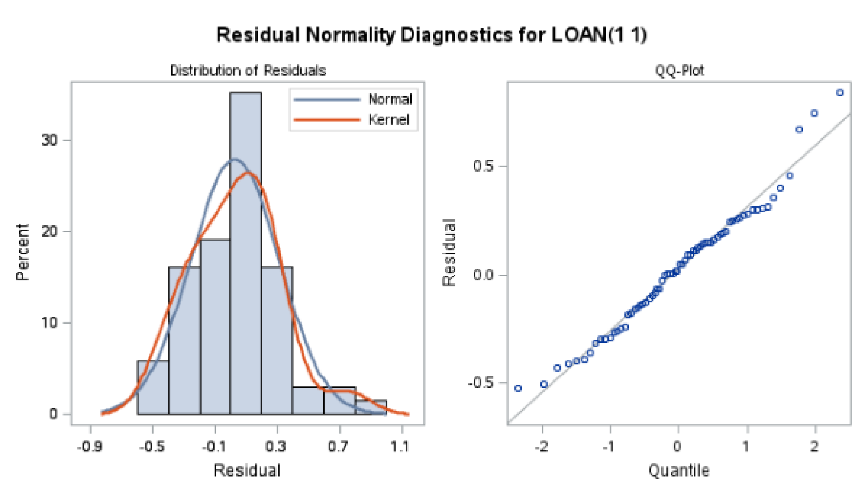


*Figure 13 Normal distribution of Residuals*

## Chapter 7: Forecasting

The model is validated as residuals show white noise behavior and hence can be used for forecasting.

| Forecasts for variable LOAN | | | | |
|---|---|---|---|---|
| Obs | Forecast | Std Error | 95% Confidence Limits | |
| 71 | 89.2664 | 0.2867 | 88.7045 | 89.8282 |
| 72 | 90.9327 | 0.5470 | 89.8606 | 92.0048 |
| 73 | 92.5991 | 0.8458 | 90.9414 | 94.2567 |
| 74 | 94.2654 | 1.1810 | 91.9508 | 96.5801 |
| 75 | 95.9318 | 1.5497 | 92.8945 | 98.9691 |
| 76 | 97.5982 | 1.9494 | 93.7775 | 101.4189 |
| 77 | 99.2645 | 2.3779 | 94.6039 | 103.9252 |
| 78 | 100.9309 | 2.8336 | 95.3772 | 106.4846 |
| 79 | 102.5972 | 3.3148 | 96.1003 | 109.0942 |
| 80 | 104.2636 | 3.8204 | 96.7757 | 111.7515 |
| 81 | 105.9300 | 4.3492 | 97.4056 | 114.4543 |
| 82 | 107.5963 | 4.9003 | 97.9919 | 117.2007 |
| 83 | 109.2627 | 5.4728 | 98.5362 | 119.9891 |
| 84 | 110.9290 | 6.0659 | 99.0401 | 122.8180 |
| 85 | 112.5954 | 6.6790 | 99.5048 | 125.6859 |
| 86 | 114.2618 | 7.3114 | 99.9318 | 128.5918 |
| 87 | 115.9281 | 7.9625 | 100.3219 | 131.5343 |
| 88 | 117.5945 | 8.6319 | 100.6763 | 134.5126 |
| 89 | 119.2608 | 9.3190 | 100.9960 | 137.5257 |
| 90 | 120.9272 | 10.0234 | 101.2818 | 140.5726 |
| 91 | 122.5936 | 10.7447 | 101.5344 | 143.6527 |
| 92 | 124.2599 | 11.4824 | 101.7548 | 146.7651 |
| 93 | 125.9263 | 12.2364 | 101.9434 | 149.9091 |
| 94 | 127.5926 | 13.0061 | 102.1012 | 153.0841 |

*Figure 14 2 Year Forecast*

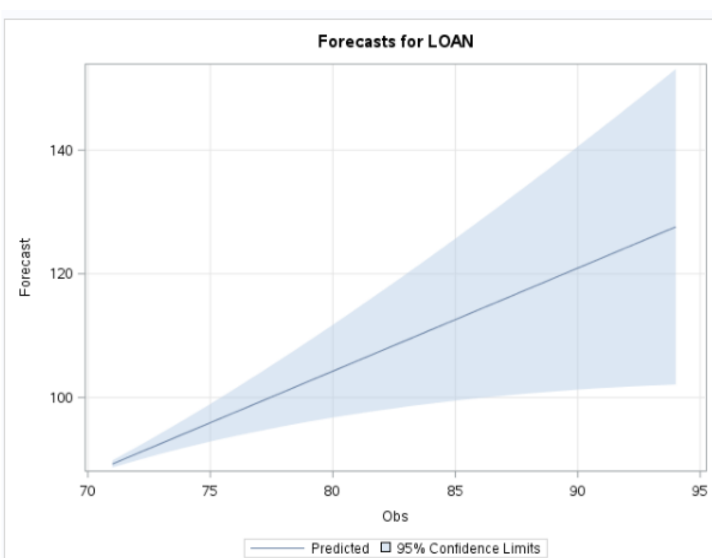


*Figure 15 Forecast with 95% confidence level*

## Result:

ARIMA (0,2,1) fits the data and forecasts is done for 2 years. Though the forecasting is done for 2 years the interval becomes too wide and hence model will compromise on accuracy if used for predicting long term future.