|  |
| --- |
| Forecasting Methods in Marketing |
| Time Series Regression Analysis  Assignment 2 |
| |  |  |  | | --- | --- | --- | | Pallavi Sethi | 7/17/20 | Forecasting Methods | |

**INDEX**

|  |  |
| --- | --- |
| **Topic** | **Page Number** |
| **Introduction** | **2** |
| **Loading Data into R** | **2** |
| **Plotting Time Series** | **3** |
| **Splitting Dataset into Two** | **4** |
| **Creating Simple Regression Model** | **5-6** |
| **Predicting and Fitting Values** | **7** |
| **Forecasting Model 1** | **8** |
| **Adding Dummy Variables** | **9** |
| **Comparing Accuracy of Models** | **10** |
| **Predicting values for new Model** | **11** |
| **Forecasting Model 3** | **12** |
| **Residual Analysis** | **13** |
| **Conclusion** | **14** |

**Introduction**

As a part of this assignment, I have taken quarterly data on GDP, Employment Rate and Fixed Mortgage Rate from 1st quarter of 1977 till last quarter of 2019. With the help of Regression analysis, I have tried to show how factors like GDP and Employment Rate influences Mortgage Rate.

**GDP**

Gross domestic product (GDP) is the total monetary or market value of all the finished goods and services produced within a country's borders in a specific time period. As a broad measure of overall domestic production, it functions as a comprehensive scorecard of a given country’s economic health.

**What is Mortgage Rate?**

A mortgage rate is the rate of interest charged on a mortgage. Mortgage rates vary for borrowers based on their credit profile. The mortgage rate is a primary consideration for homebuyers looking to finance a new home purchase with a mortgage loan.

Factors like gross domestic product (GDP) and the employment rate, can influence mortgage rates. Higher economic growth levels generally produce higher incomes and higher levels of consumer spending, including more consumers seeking mortgage loans for home purchases. Other factors like Housing market conditions also affect mortgage rates. When fewer homes are being built or offered for resale, the decline in home purchasing leads to a decline in the demand for mortgages and pressures interest rates downward.

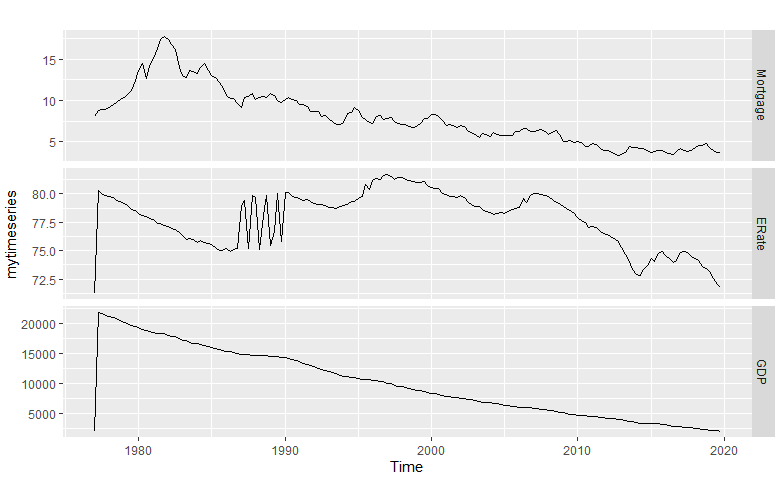
***Loading Data***

In order to load the data into R environment run and for setting the working directory run ***steps 1-3*** from the appendix.

For doing time series analysis we need to convert the csv data into ts() object. Run ***Step 4,*** also check the class of ‘mytimeseries’ using Class() and str() function to check the start and end time of the data.

***Plotting Time Series***

Once the time series data(mytimeseries) is loaded into R, the next step is to make a plot of the time series for better comparison. Run ***step 5*** from the appendix

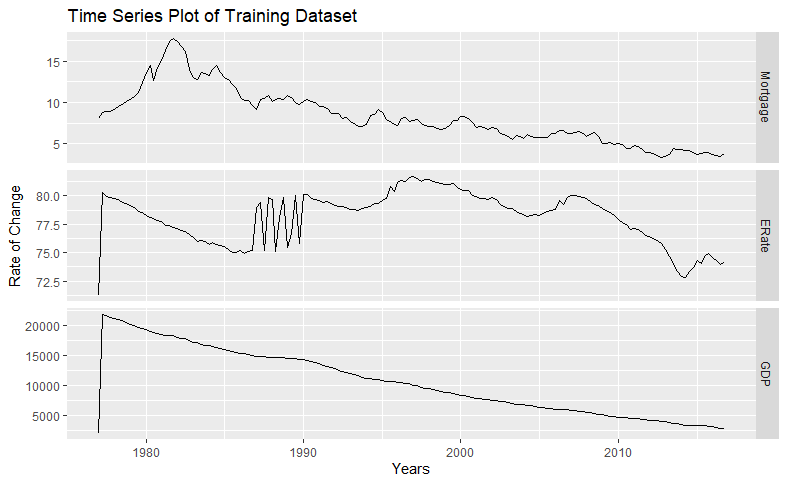
******

***Findings from the Time Series:***

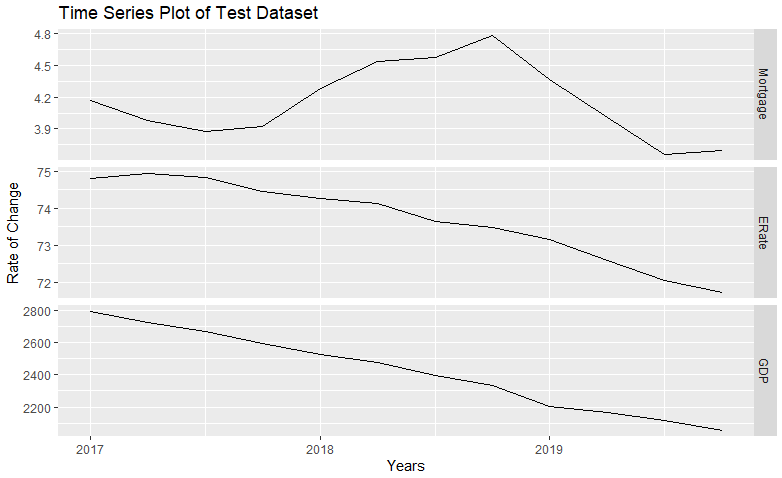
Mortgage rate were all time higher during the 1980’s, since then we can see the rate declines gradually over time. Factors like Employment Rate and GDP was also decreased during that period.

***Splitting the data into training and test sets***

The complete dataset is broken into 2 datasets: Training Data and Testing Data; run ***step 6*** from the appendix to divide the data and view time series plot for both the datasets.

*Training Data is depicted below*

*Time Series plot for Test Dataset*



***Creating simple time series regression model with just one predictor and one forecast variable using the training data.***

Run ***step 7*** from the appendix to perform simple regression using tslm() on predictor(‘GDP’) and forecast variable(‘Mortgage Rate’).

Run ***Step 8*** for summary results and detailed regression results of tslm().

I got the below summary of the model

Call:

tslm(formula = Mortgage ~ GDP, data = model1)

Residuals:

Min 1Q Median 3Q Max

-5.3934 -0.5655 -0.1674 0.4928 5.4157

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.5695245 0.2875233 8.937 9.94e-16 \*\*\*

GDP 0.0005354 0.0000242 22.123 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.662 on 158 degrees of freedom

Multiple R-squared: 0.756, Adjusted R-squared: 0.7544

F-statistic: 489.4 on 1 and 158 DF, p-value: < 2.2e-16

Call:

tslm(formula = Mortgage ~ GDP, data = model1)

Residuals:

Min 1Q Median 3Q Max

-5.3934 -0.5655 -0.1674 0.4928 5.4157

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.5695245 0.2875233 8.937 9.94e-16 \*\*\*

GDP 0.0005354 0.0000242 22.123 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.662 on 158 degrees of freedom

Multiple R-squared: 0.756, Adjusted R-squared: 0.7544

F-statistic: 489.4 on 1 and 158 DF, p-value: < 2.2e-16

From the above result we can say that the Mortgage Rate and GDP have a positive relationship with a positive slope of 0.0005, with adjusted R² is 75% that means this Model predicts the value of Mortgage with 75% accuracy.

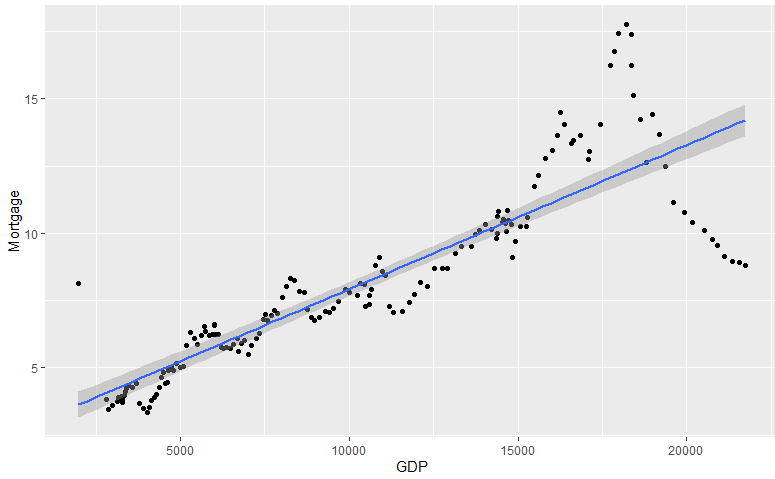
P value is less that alpha(0.05).

Regression equation for these 2 variables would be

**Mortgage = 2.56 + 0.0005(GDP)**

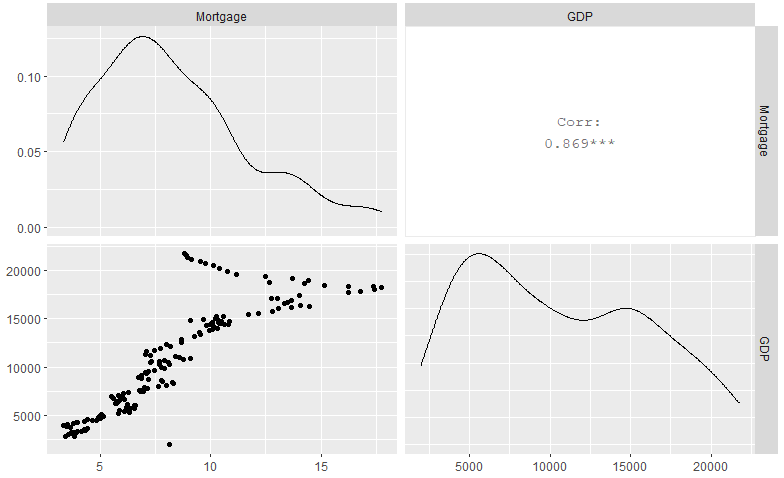
From the above equation we can say that 1 unit increase in GDP would result in 0.005x in Mortgage.

***Plotting Regression result***

******

**From the regression plot we can observe that** not all the data points lie on the straight line but are scattered around it, Infact it shows cyclical pattern which might be because of Correlation between the variable.

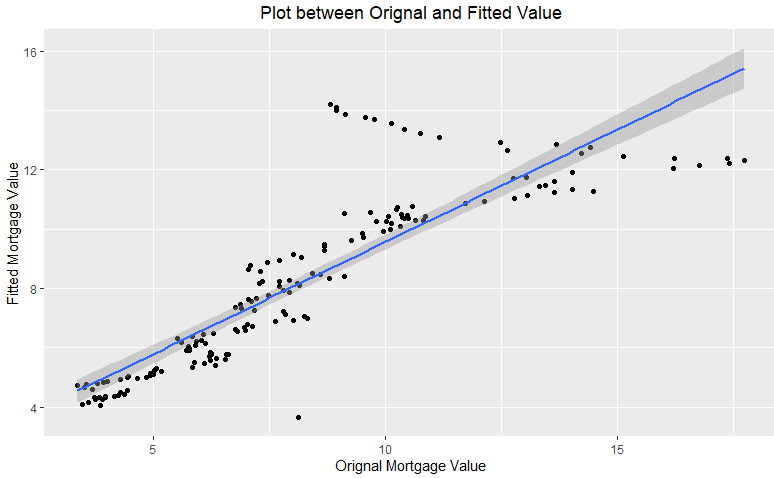
Run Step 8 of appendix, in order to check correlation between the GDP and Mortgage.

From the graph it is very clear that both the variables are strongly correlated, but this doesn’t necessary meant causation.

***Predicting values and fit in the model***

Fitted values are the  predictions of the data used to estimate the model, not genuine forecasts of future of forecast variable.

Run ***step 9.*** to calculate the fitted value of out forecast variable using fitted(). After calculating the fitted value, we plot the graph between original and fitted value.



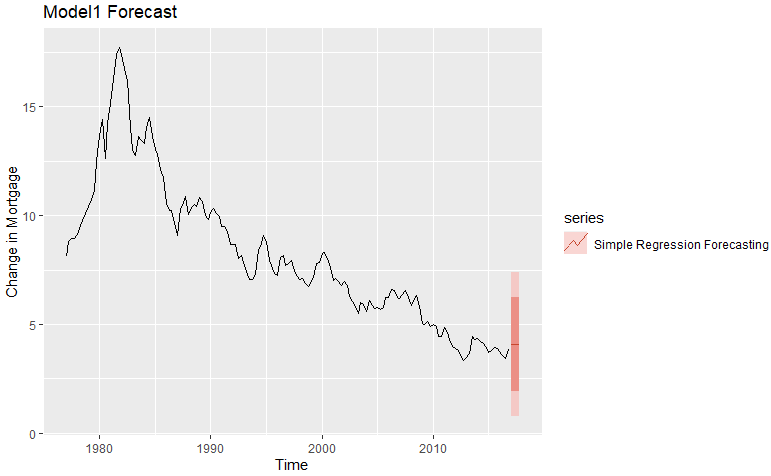
We can see that majority of the data points lies near the line. That means the original and fitted values obtained from R are very close, except for few points which lies far from the line.

**Forecasting the Model**

The next in the analysis is to forecast the model we created for the future values of Mortgage. Run step10 from the appendix.

I have used forecast() on Model1. In this forecast I have kept the value of GDP as constant(the last occurred value of GDP) and then forecast the value of Mortgage for all 4 quarters of next year.

The dark orange shading represents the range of low and high values for 80% confidence, the light color shows the range for 95% confidence Interval.



So far, we have seen the regression analysis and forecasting, on 1 predictor and 1 forecast variable. But sometime underlying trend and season also impact the forecast. These variables are called Dummy Variables.

A dummy variable can also be used to account for an **outlier** in the data. Rather than omit the outlier, a dummy variable removes its effect.

For the remaining models, I have added dummy variables(model2) and Employement Rate(model3) to my existing model1 to see if these variables have a strong relation with the forecast variable and ultimately generate a more robust forecast model.

Run step 11 from the appendix to create model2 and model3

Model2 have added trend and season to Model1 which currently shows the effect of GDP on Mortgage, whereas in Model 3 I have added ERate(Employment Rate) along with GDP to see if ERate also influence Mortgage.

**Summary result of Model 2**

Response Mortgage :

Call:

tslm(formula = Mortgage ~ trend + season)

Residuals:

Min 1Q Median 3Q Max

-5.0202 -0.6581 -0.1143 0.5090 5.6329

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 13.203919 0.335686 39.334 <2e-16 \*\*\*

trend -0.063679 0.002762 -23.054 <2e-16 \*\*\*

season2 0.207679 0.360752 0.576 0.566

season3 0.215359 0.360784 0.597 0.551

season4 0.176788 0.360837 0.490 0.625

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.613 on 155 degrees of freedom

Multiple R-squared: 0.7743, Adjusted R-squared: 0.7685

F-statistic: 132.9 on 4 and 155 DF, p-value: < 2.2e-16

From the summary we can say that this model is not very much effective in highlighting he relationship between the variables because of the high P value.

Thus, we can conclude that the model with dummy variables is not the better model.

I have also compared the accuracy of Model 1 and Model2.

accuracy(model2)

ME RMSE MAE MPE MAPE MASE ACF1

Training set -1.483702e-13 1095.528 349.5453 -2.798 12.22286 0.6912878 0.1028508

> accuracy(model1)

ME RMSE MAE MPE MAPE MASE ACF1

Training set 1.553441e-16 1.651138 1.036332 -3.010188 11.85965 1.153781 0.8186739

It is said that model with lower RMSE is a better model, thus model 1 is better than model2.

**Summary result of model 3**

summary(model3)

Call:

tslm(formula = Mortgage ~ GDP + ERate, data = model1Data)

Residuals:

Min 1Q Median 3Q Max

-5.1440 -0.8253 -0.0160 0.7021 5.1778

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.595e+01 4.495e+00 3.549 0.000511 \*\*\*

GDP 5.467e-04 2.392e-05 22.855 < 2e-16 \*\*\*

ERate -1.728e-01 5.794e-02 -2.983 0.003313 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.622 on 157 degrees of freedom

Multiple R-squared: 0.769, Adjusted R-squared: 0.7661

F-statistic: 261.4 on 2 and 157 DF, p-value: < 2.2e-16

accuracy(model3)

ME RMSE MAE MPE MAPE MASE ACF1

Training set -1.330736e-16 1.606253 1.088474 -3.438116 13.69081 1.211832 0.8434175

accuracy(model1)

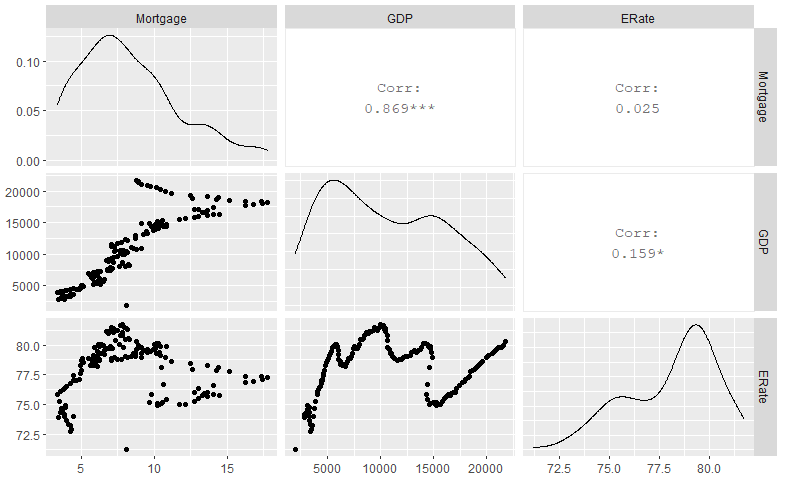
ME RMSE MAE MPE MAPE MASE ACF1

Training set 1.553441e-16 1.651138 1.036332 -3.010188 11.85965 1.153781 0.8186739

From the above Summary result of model 3, we can conclude that Model 3 better than model1 and model2. The adjusted R² for model 3 is 76.61% which is greater than base model 1 and the P value is less than 0.05 (alpha), thus we can say that adding ERate and GDP to the model, we would get a better model that explains the variation of Mortgage Rate. The RMSE of model3 is also lower that model 1, thus it is more accurate.

**Mortgage = 15.9514756016 + 0.00054(GDP) -0.172 (ERate)**

**Checking Correlation between variables in Model 3**

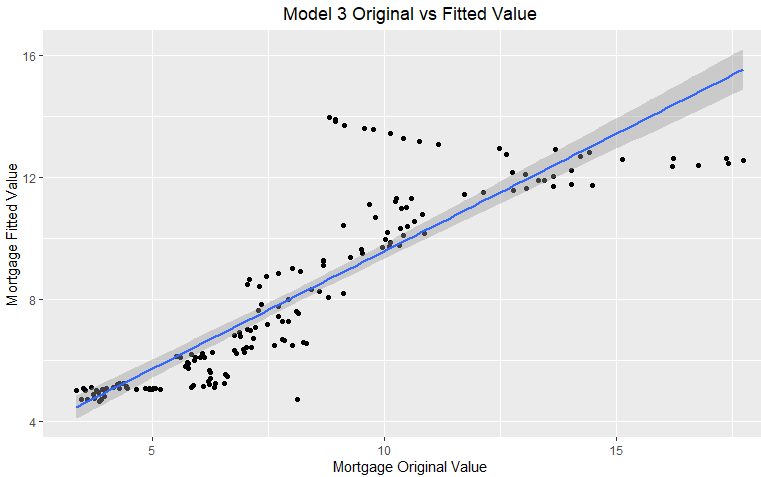
****

From the above graph, we can say that there is positive correlation between GDP, ERate and Mortgage. The patterns appear in the plots above might be because of the existing correlation between the variables.

**Predicting Values and fit for Model 3**

Run ***step 12*** for calculating the fitted valued for the variables in model 3.

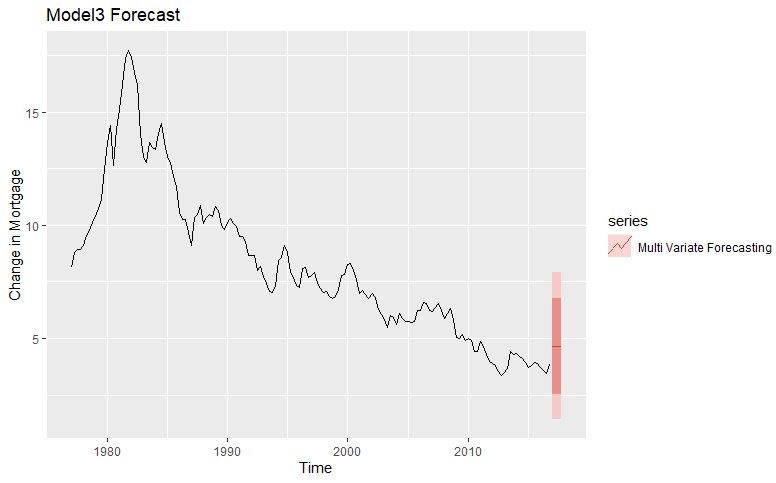
Below graph highlighted the values of Original and Fitted value of Mortgage in Model 3

****

The predicted values of Mortgage in Model 1 and Model 3 are very close.

**Forecasting prediction for model 3**

Run ***step 13*** from the appendix

****

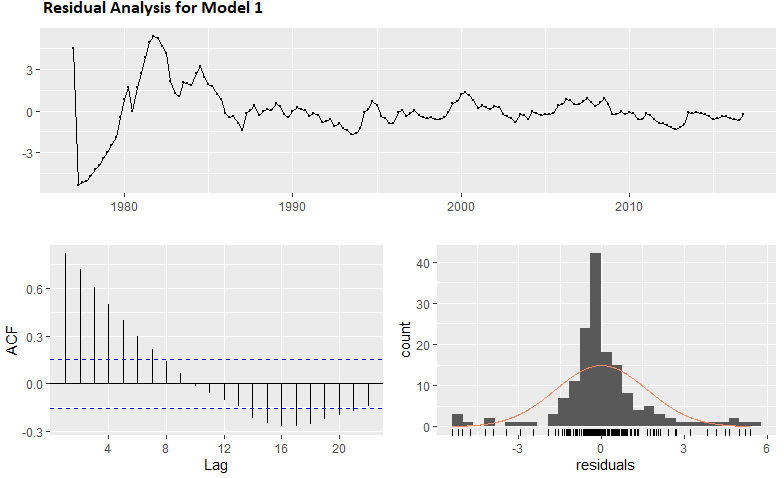
While calculating the forecast of Mortgage in Model 3, I have kept the values of GDP and ERate constant (last reported value). The darker region showed the range of Mortgage for 80% confidence interval whereas the lighter region showed the range with 95% confidence interval.

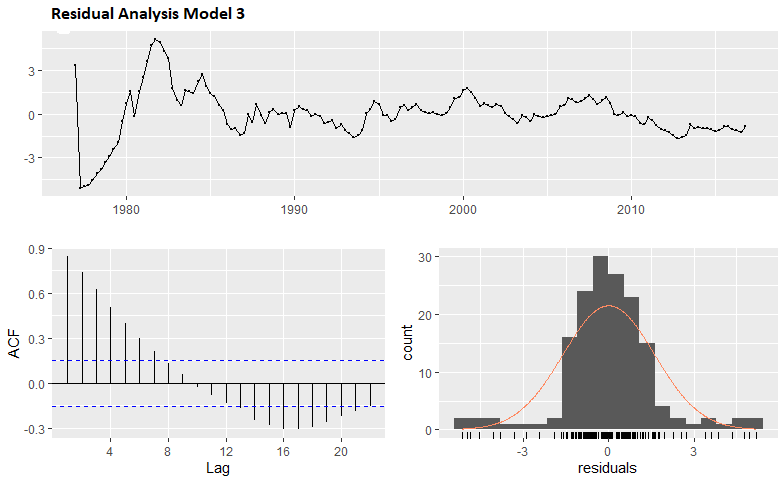
**The last step in this analysis is checking the residuals**

Residuals are useful in checking whether a model has adequately captured the information in the data. A good forecasting method will yield residuals with the following properties:

1. The residuals are uncorrelated. If there are correlations between residuals, then there is information left in the residuals which should be used in computing forecasts.
2. The residuals have zero mean. If the residuals have a mean other than zero, then the forecasts are biased.
3. The residuals have constant variance.
4. The residuals are normally distributed.

Run ***Step 15,***  for checking the residuals in each of our models.





From the above residual plot, we can say that, for both the models the mean of the residuals is closer to zero and the residual shows constant variance except at one part(during 1980 Mortgage surge cycle)

**Conclusion**:

Mortgage rates are tied to the basic rules of supply and demand. Interest rates on residential mortgages can be influenced by longer-term trend changes of various economic indicators like GDP and Employment Rate. Thus, understanding of key economic indicators can provide clues to the future direction of interest rates.