A time series analysis of GDP of Nepal

Using AMIRA model

This is an example of time series analysis of GDP of Nepal using ARIMA (Autoregressive integrated moving average model) using R programming. The goal of the project is to predict the GDP of Nepal for next 10 years. This project was done as a term project for the course STA 614: research methodology @ Central department of statistics, Tribhuvan University.

Packages required

* library (rvest)
* library (dplyr)
* library (ggplot2)
* library (forecast)
* library(tseries)

All downloadable from CRAN repositories.

Prerequisites

* Knowledge of R programming
* R studio

Data description

The data was scrapped from the link [*https://www.macrotrends.net/countries/NPL/nepal/gdp-gross-domestic-product*](https://www.macrotrends.net/countries/NPL/nepal/gdp-gross-domestic-product)*.* The data gives us the GDP (nominal) of Nepal from 1960 to 2020 along with GDP per capita and percent growth in GDP.

Visualization of GDP

The ggplot2 is an excellent library in R for interactive and appealing visualization. The following chat was generated by ggplot2 and show the GDP of Nepal from 1960 to 2020.

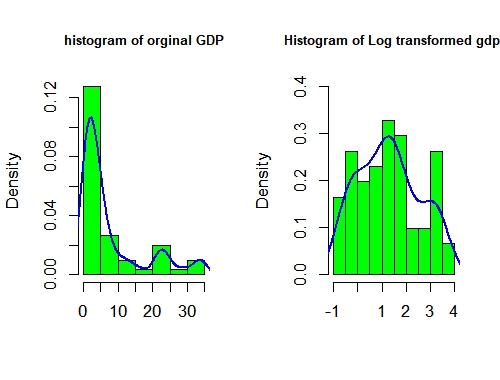


*Figure 1: Line chart of GDP of Nepal from 1960 to 2020*

Prediction Model Description:

The goal of this project is to predict the future GDP of Nepal using ARIMA model. For the implementation of ARIMA model the time series must have to be stationary. Stationary can be obtained be differencing the series for suitable number of time ,but as the histogram of original data revels that the data is positively skewed, so we first take the log transformation to make the data more normal. Then stationarity is obtained by differencing the log-transformed series.

To check the stationary we had used a celebrated augmented dicky Fuller test of stationary along with ACF (autocorrelation function) and PACF (partial autocorrelation function plot) to cross check.



*Figure 2: Histogram of original vs log transformed data*

Finding the best ARIMA model for the transformed series:

After the series has been log-transformed, next task is to find the best ARIMA model for the series. An ARIMA model consists of three term p, d and q.

p: Number of autoregressive terms

d: number of time the series should be differenced to make it stationary

q: Number of moving average terms

#Result of ADF (augmented dickey fuller) test of stationary

*Augmented Dickey-Fuller Test*

*data: log\_transfromed\_series*

*Dickey-Fuller = -1.4818, Lag order = 3, p-value = 0.7844*

*alternative hypothesis: stationary*

*Augmented Dickey-Fuller Test*

*data: diff\_series*

*Dickey-Fuller = -4.7628, Lag order = 3, p-value = 0.01*

*alternative hypothesis: stationary*

Results from augmented Dickey-Fuller test shows that the first differenced series of log transformed series is stationary. Now the best ARIMA model for the given log transformed series is obtained by using *auto.arima ()* function in the FORECAST R package which gives the best ARIMA model based on the AICc (Akaike information criteria corrected for sample size).The result is shown below:

*ARIMA (2, 1, 2) with drift : -103.9589*

*ARIMA (0, 1, 0) with drift : -103.3401*

*ARIMA (1, 1, 0) with drift : -101.408*

*ARIMA (0, 1, 1) with drift : -102.196*

*ARIMA (0, 1, 0) : -81.15208*

*ARIMA (1, 1, 2) with drift : -106.3378*

*ARIMA (0, 1, 2) with drift : -108.7183*

*ARIMA (0, 1, 3) with drift : -106.3382*

*ARIMA (1, 1, 1) with drift : -104.722*

*ARIMA (1, 1, 3) with drift : -104.1381*

*ARIMA (0, 1, 2) : -88.63073*

*Best model: ARIMA (0, 1, 2) with drift*

*Series: gdp\_tseries*

*ARIMA (0, 1, 2) with drift*

*Box Cox transformation: lambda= 0*

*Coefficients:*

*ma1 ma2 drift*

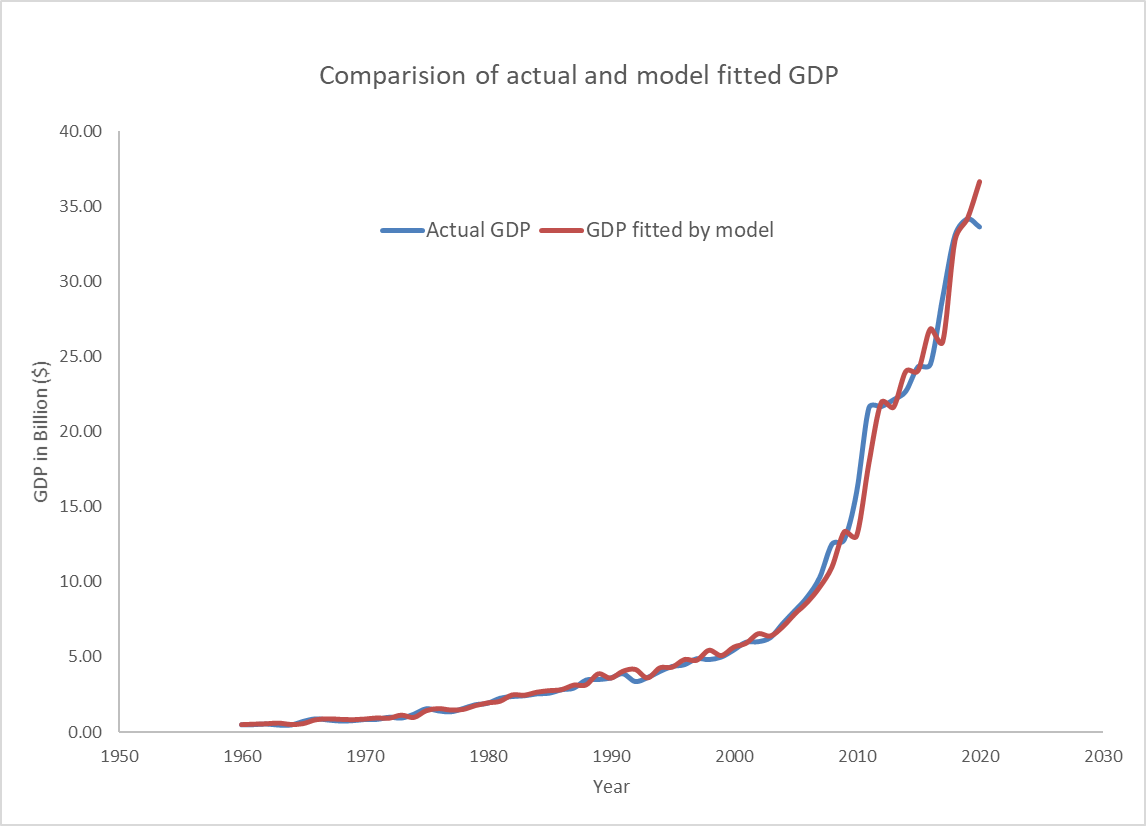
*0.1172 -0.4004 0.0706*

*s.e. 0.1233 0.1251 0.0085*

*Sigma^2 = 0.008648: log likelihood = 58.72*

*AIC=-109.45 AICc=-108.72 BIC=-101.07*

The result from R gives ARIMA (0, 1, 2) with drift is a best model with minimum AICc of

-108.72for our data.

*Figure 3: Comparison of Actual GDP vs model fitted GDP*

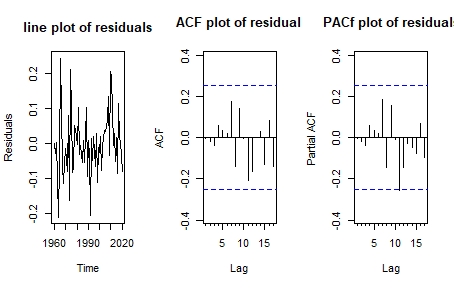
For further checking the adequacy of our model we will check the residual for its stationary. For this we will apply ADF test for residuals and check the ACF and PACF of residuals. The results is shown below:

*Augmented Dickey-Fuller Test*

*data: residuals*

*Dickey-Fuller = -3.681, Lag order = 3, p-value = 0.03387*

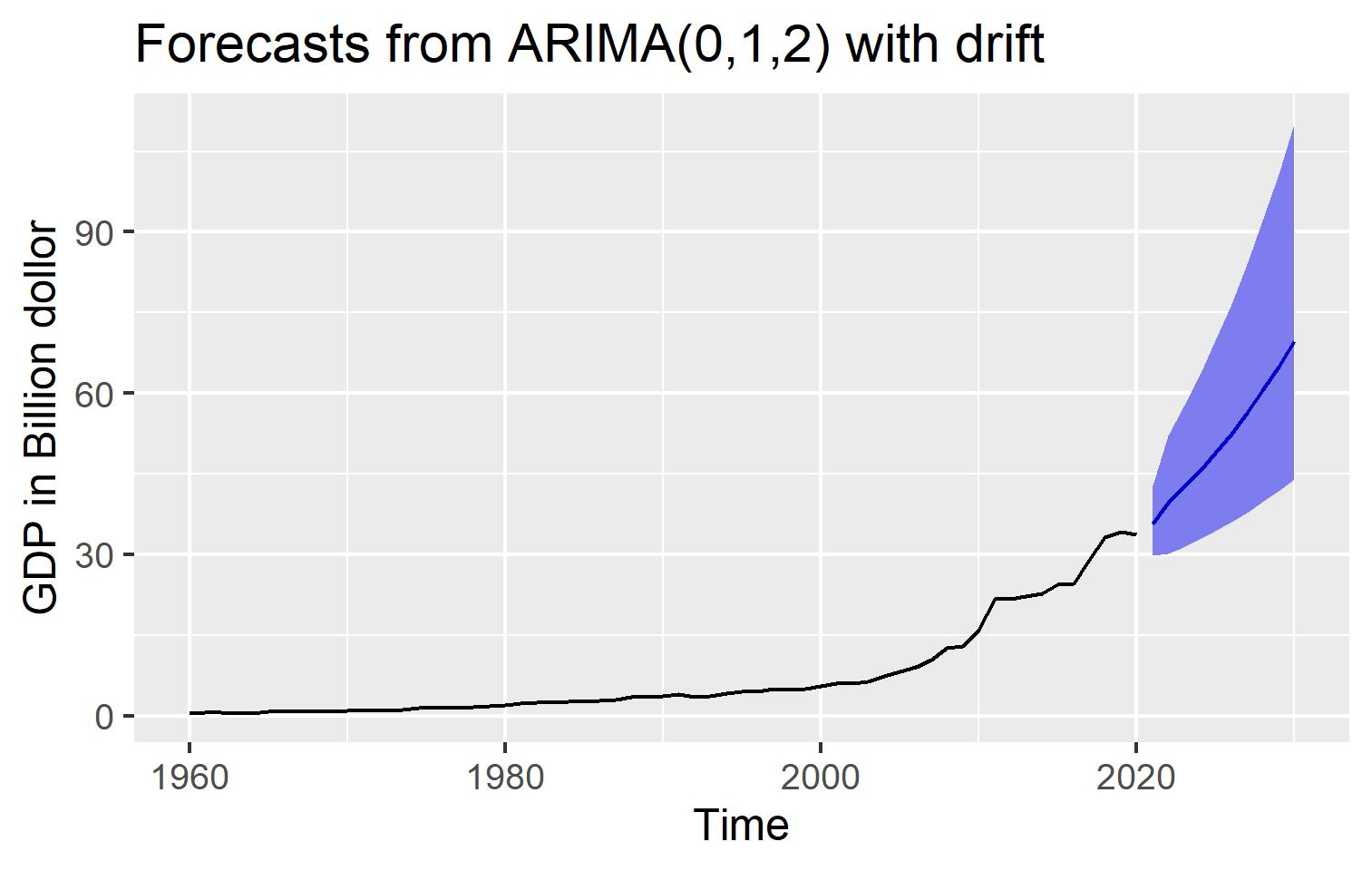
*alternative hypothesis: stationary*



*Figure 4: Analysis of residuals from the fitted model*

The residual analysis shows that residual from our fitted model is stationary. Hence the model is appropriate for our data. Now based on the fitted model we have predicated the GDP of Nepal up to 2030. The predicted values is shown in a table below:

|  |  |  |  |
| --- | --- | --- | --- |
| **year** | **predicted** | **95%upperbound** | **95%lowerbound** |
| 2021 | 35.71116 | 42.8508 | 29.7611 |
| 2022 | 39.58005 | 52.0184 | 30.11589 |
| 2023 | 42.47382 | 57.49966 | 31.37454 |
| 2024 | 45.57916 | 63.39036 | 32.77249 |
| 2025 | 48.91154 | 69.74183 | 34.30278 |
| 2026 | 52.48755 | 76.60398 | 35.96344 |
| 2027 | 56.32501 | 84.02752 | 37.75557 |
| 2028 | 60.44304 | 92.06531 | 39.68227 |
| 2029 | 64.86214 | 100.7732 | 41.74819 |
| 2030 | 69.60433 | 110.2105 | 43.95916 |

Finally the Forecasted values are plotted:

*Figure 5: Forecasted GDP using the selected model*