# **GDP growth predict project**

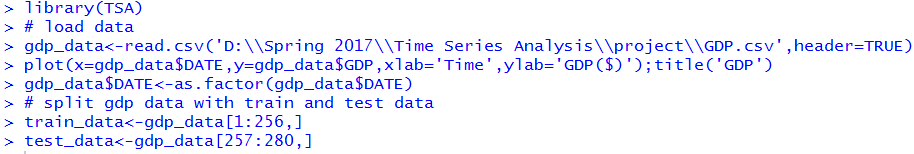
1. Data Source:

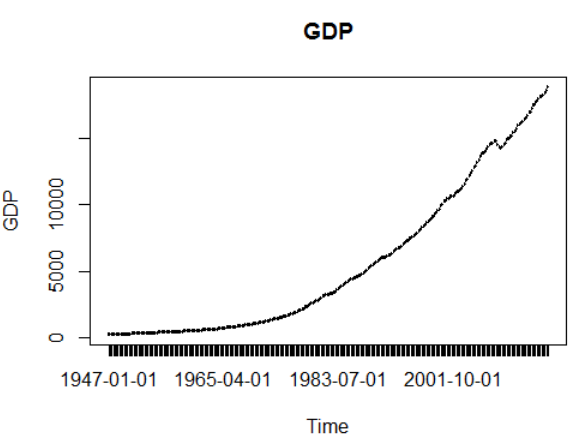
Gross Domestic Product (GDP) is a monetary measure of the market value of all final goods and services produced in a period (quarterly or yearly). Nominal GDP estimates are commonly used to determine the economic performance of a whole country or region, and to make international comparisons. The Federal Reserve Bank in Saint Louis, Missouri has an economic database, called FRED, containing 41,000 economic series, available at no cost via the following portal which includes GDP data.

<http://research.stlouisfed.org/fred2>

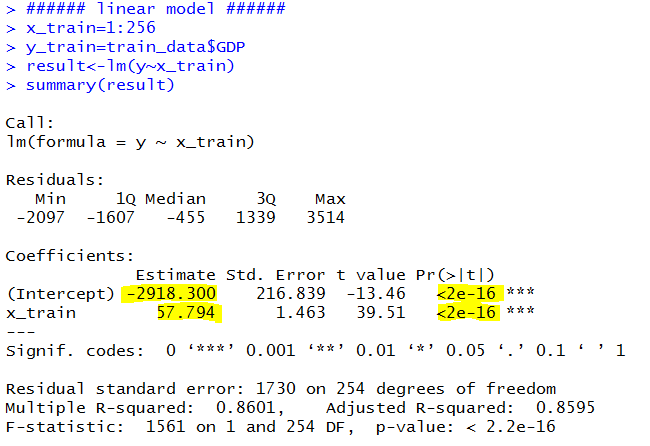
The GDP data from this portal in a file called GDP.csv . This file contains GDP data from January 1, 1947 (1Q 1947) till October 1, 2016 (4Q 2016) on a quarterly basis.

1. Load data and split dataset with train data and test data



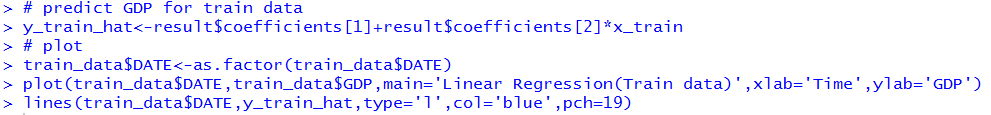


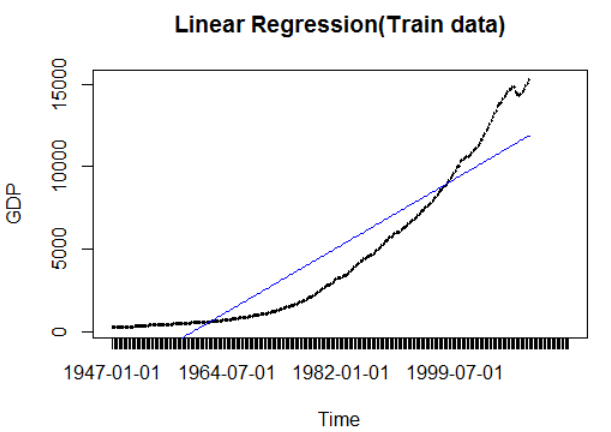
1. Linear forecasting mode
   1. Build up the model with train dataset



The model is y=57.794\*x-2918.300

* 1. Predict for training data

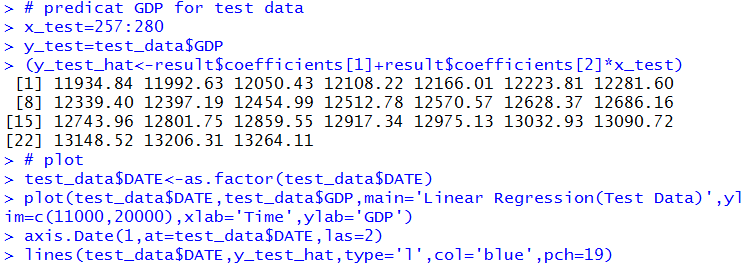


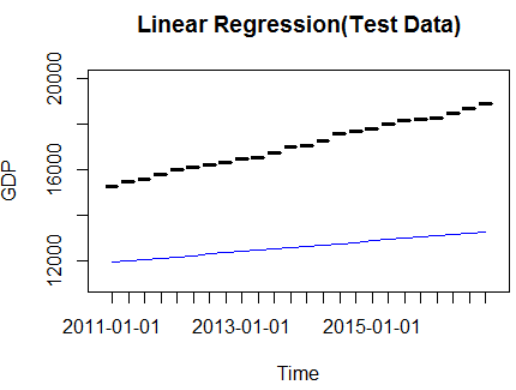


* 1. Compute RMSE for training data



* 1. Predict for test data

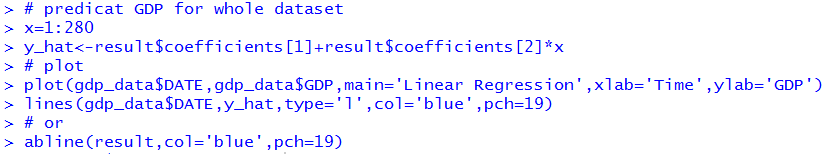


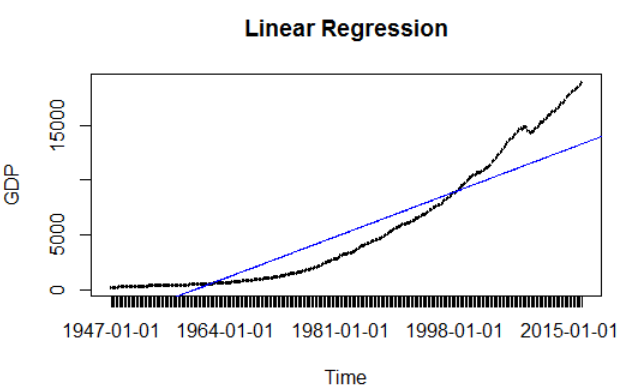


* 1. Compute RMSE for test data

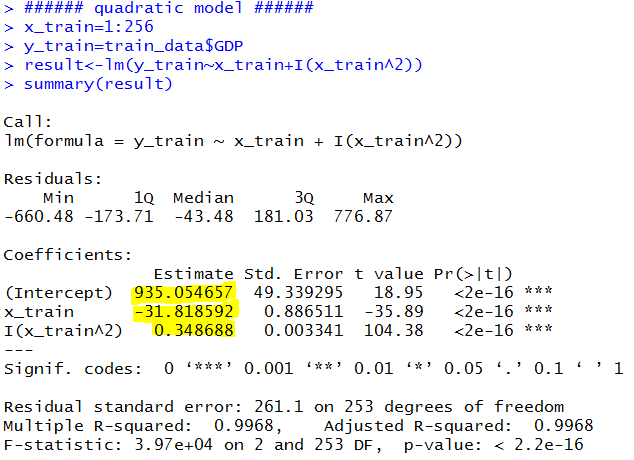


* 1. Plot and predict the whole dataset





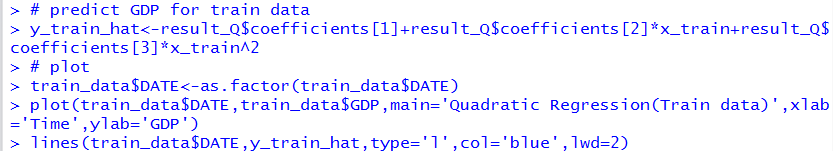
1. Quadratic trend forecasting model
   1. Build up the model with train dataset

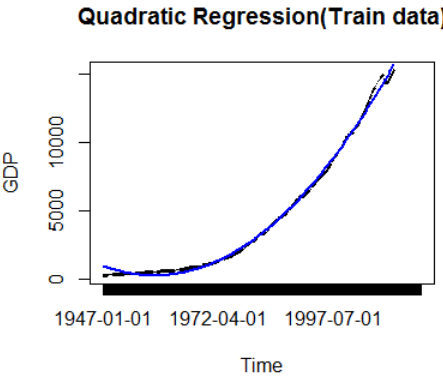


The model is

y=0.348688\*x2-31.818592\*x+935.054657

* 1. Predict for training data

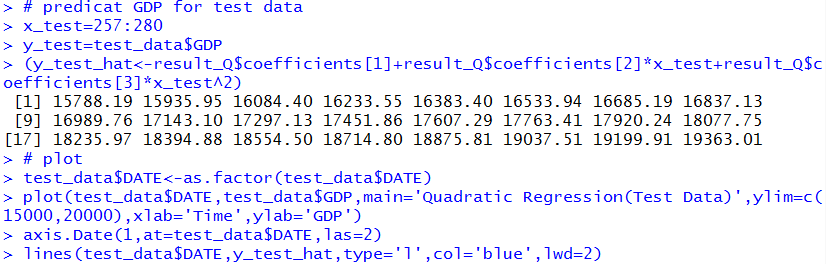


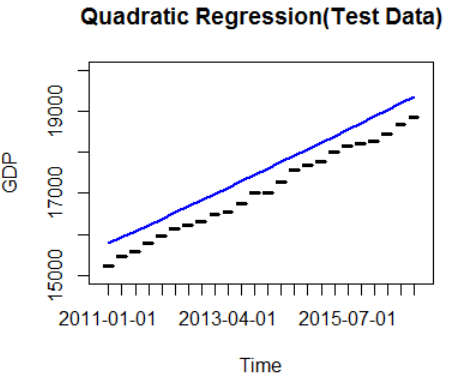


* 1. Compute RMSE for training data



* 1. Predict for test data

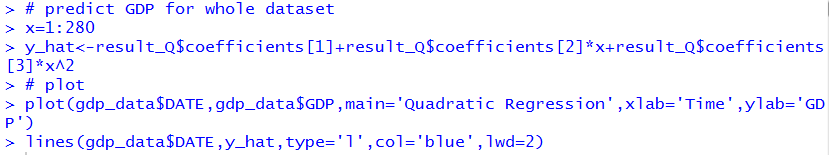


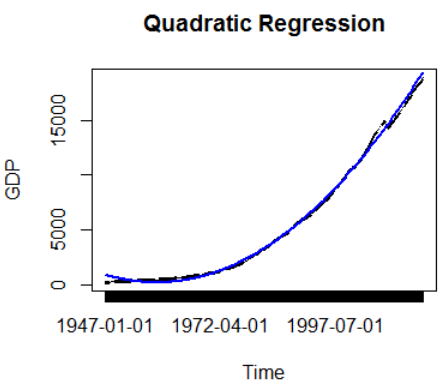


* 1. Compute RMSE for test data

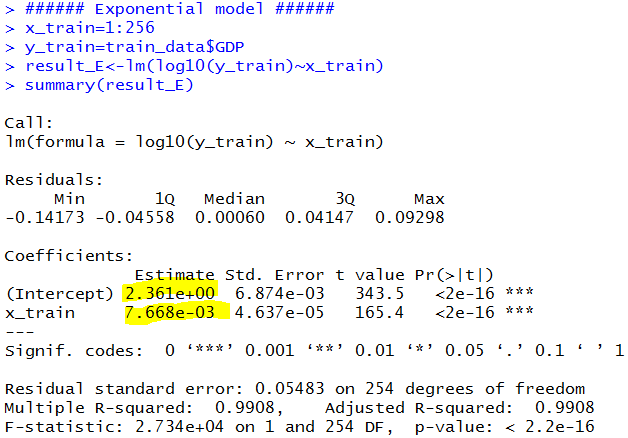


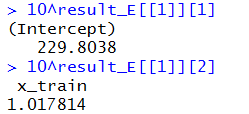
* 1. Plot and predict the whole dataset





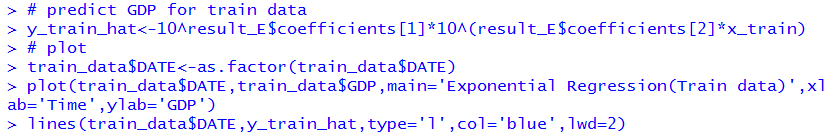
1. Exponential trend forecasting model
   1. Build up the model with train dataset

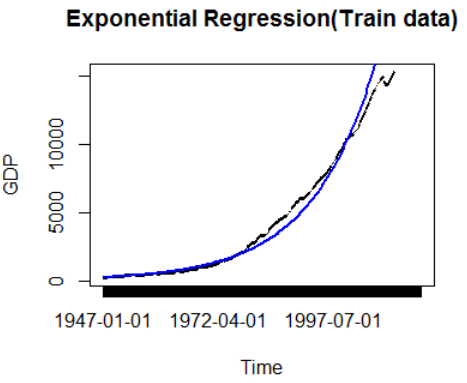




The model is y=1.017814x \* 229.8038

* 1. Predict for training data

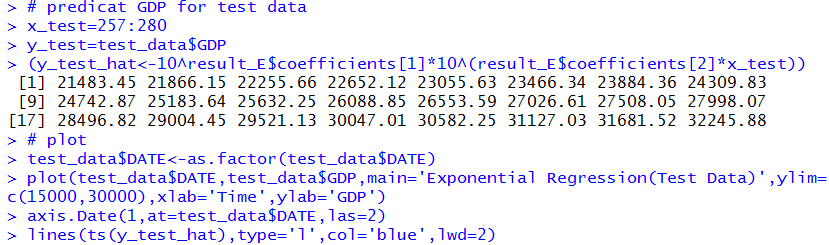


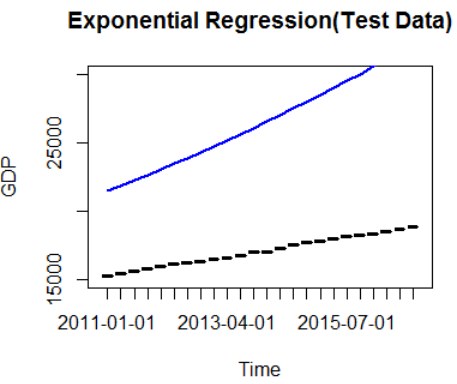


* 1. Compute RMSE for training data



* 1. Predict for test data

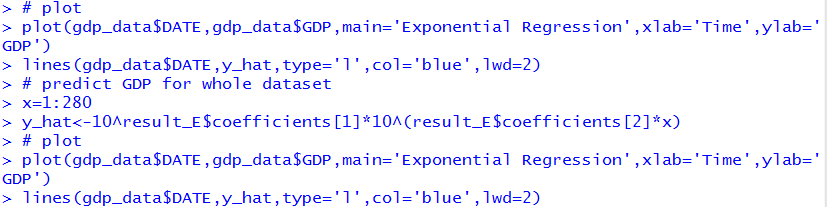


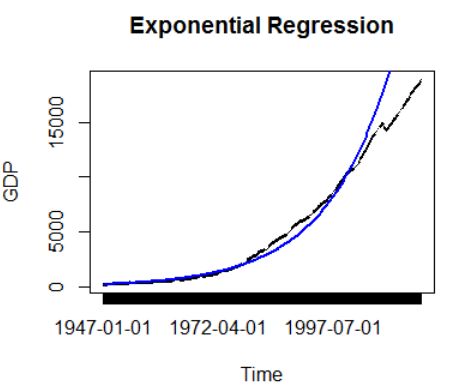


* 1. Compute RMSE for test data

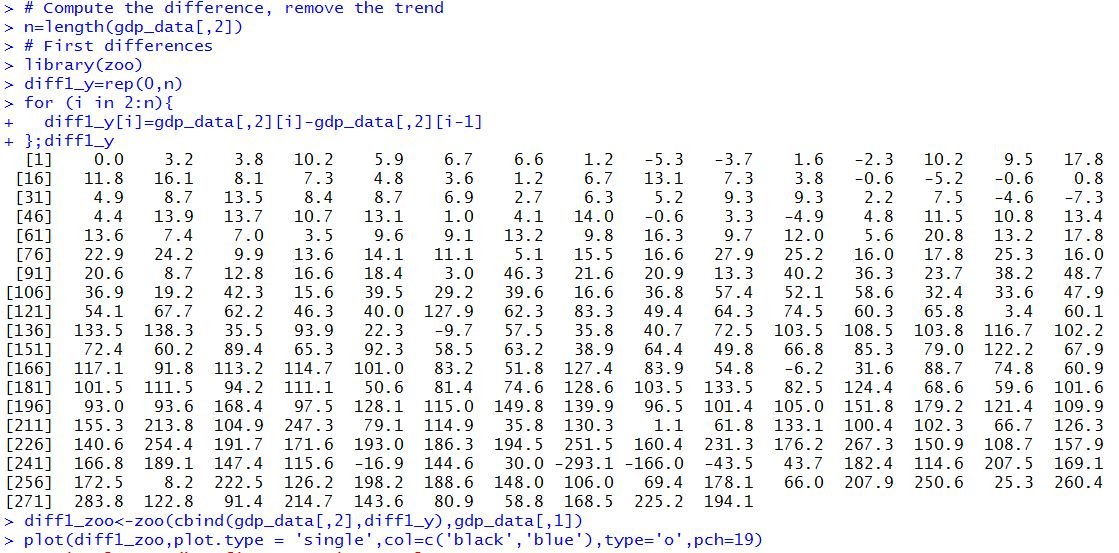


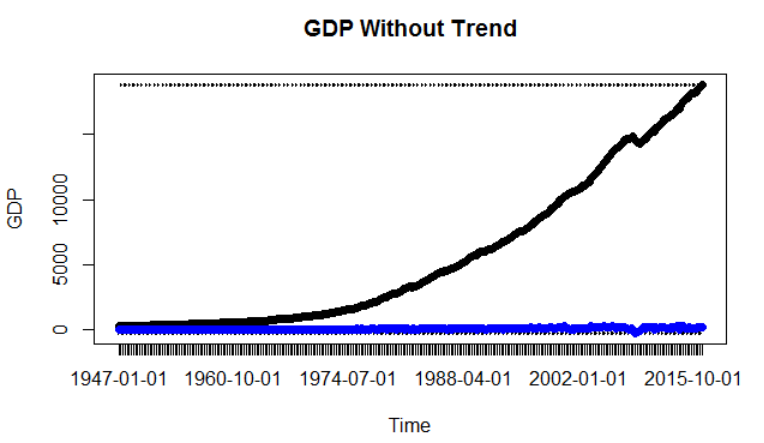
* 1. Plot and predict the whole dataset



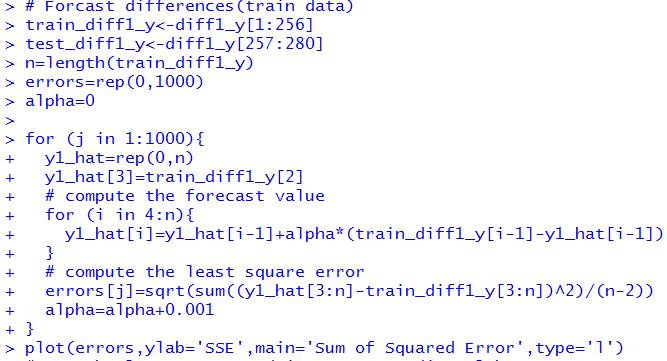


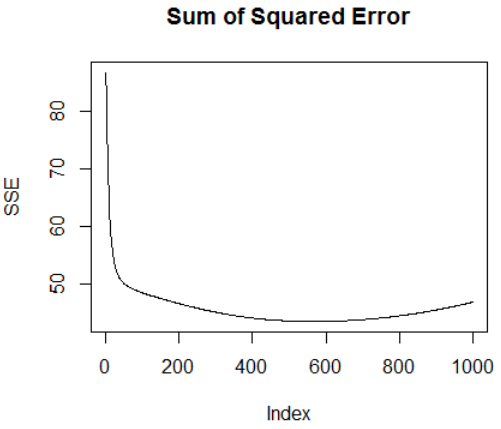
1. Exponential Weighted Moving Average model
   1. Compute the difference of the series to remove the trend.



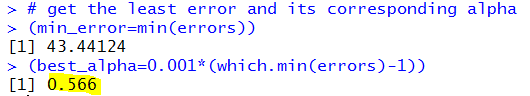


* 1. Vary the ‘alpha’ value from 0 to 1 with an increment of 0.001 and compute SSE(Sum of Squared Error).





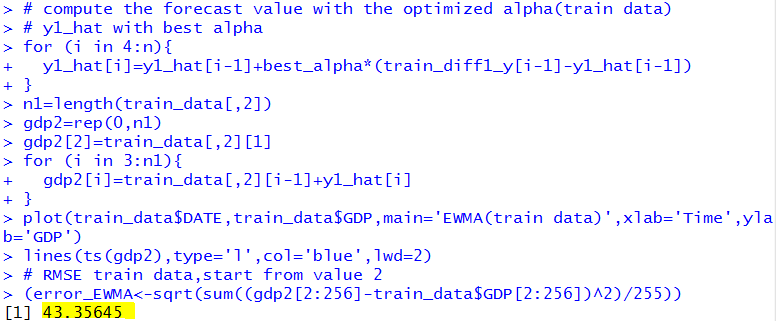
* 1. Find the optimum value of ‘alpha’ for which the SSE is least.

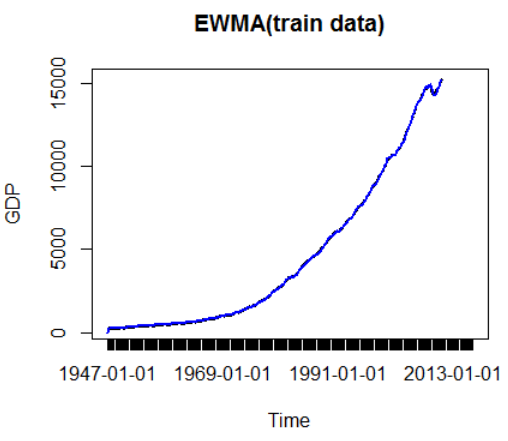


**When alpha=0.566, we get the least SSE=43.44124**

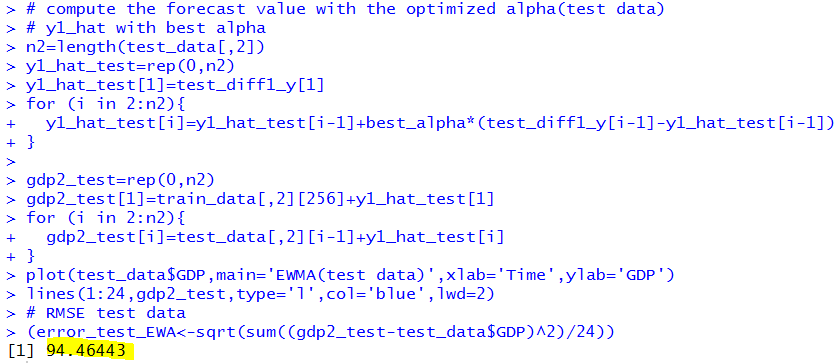
* 1. By using this optimum value of ‘alpha’, compute the forecast and its RMSE.

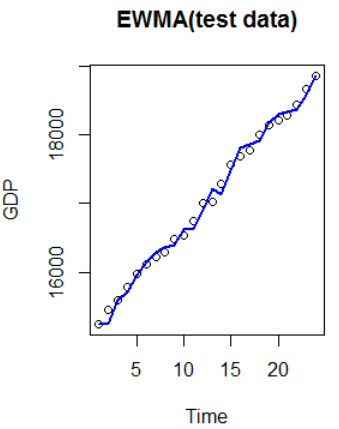
**Training dataset:**



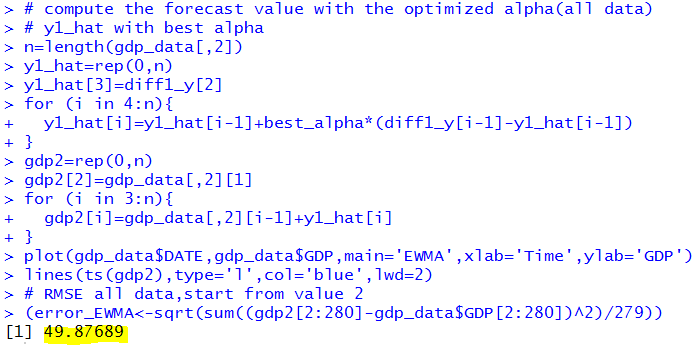


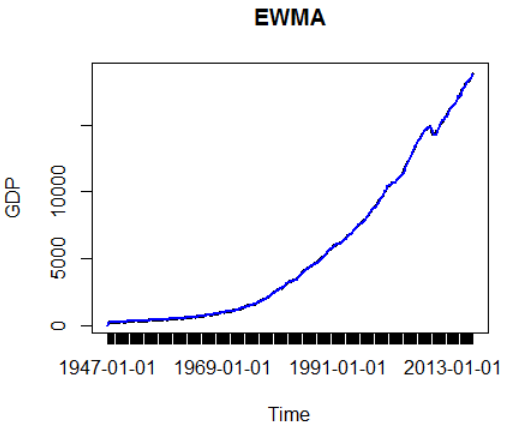
Test data:



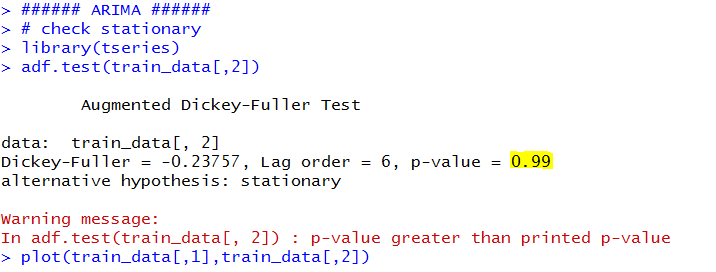


Whole dataset:

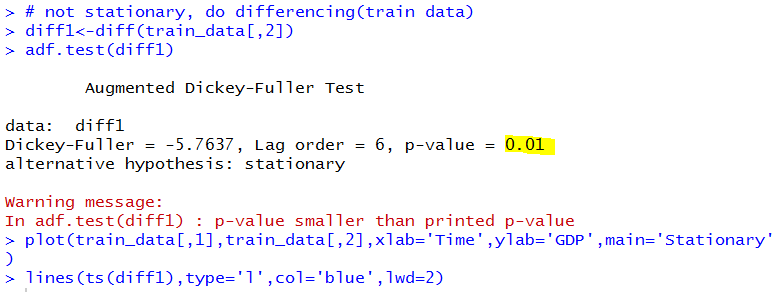


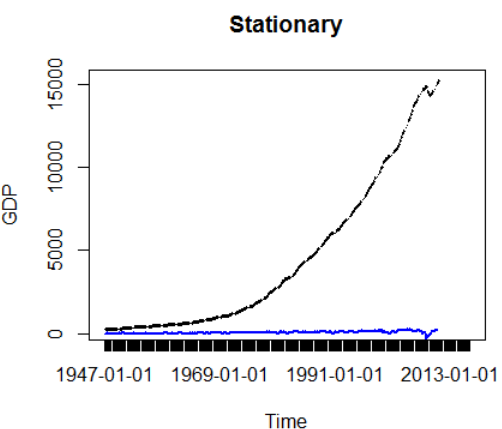


1. ARIMA model
   1. Test this dataset for stationarity. If the series is not stationary, make it stationary.



**p-value = 0.99 > 0.05, it’s not stationary.**

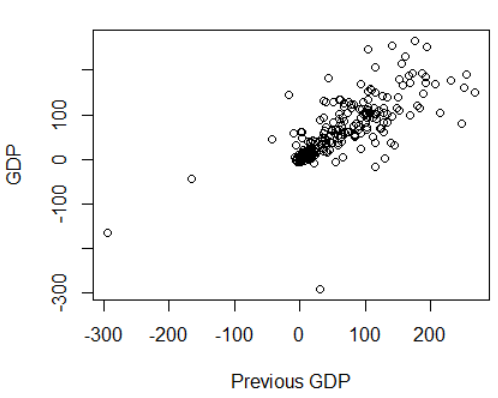




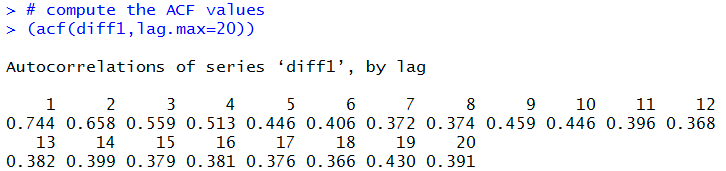
**After the first difference, p-value=0.01<0.05, it becomes stationary.**

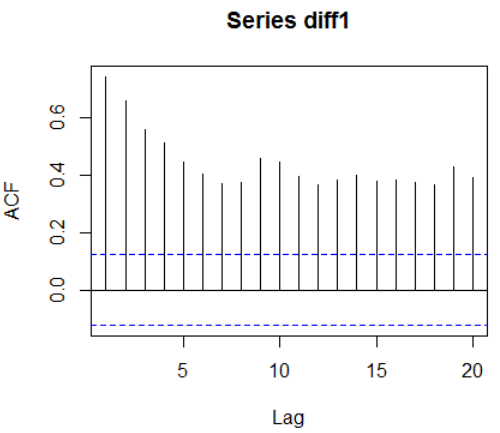
* 1. Compute the ACF values



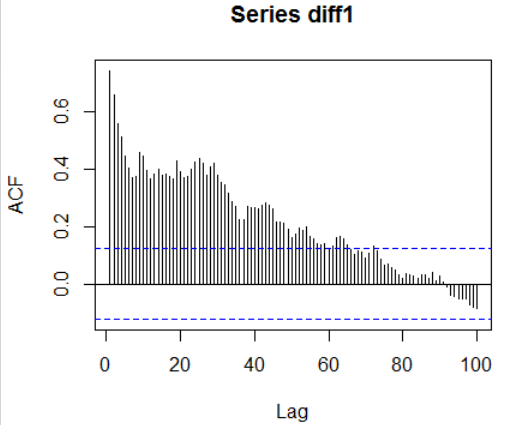


**We can see there GDP is dependent on the previous GDP.**



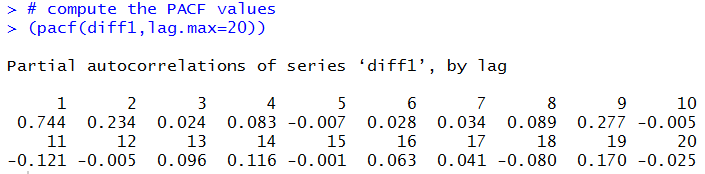


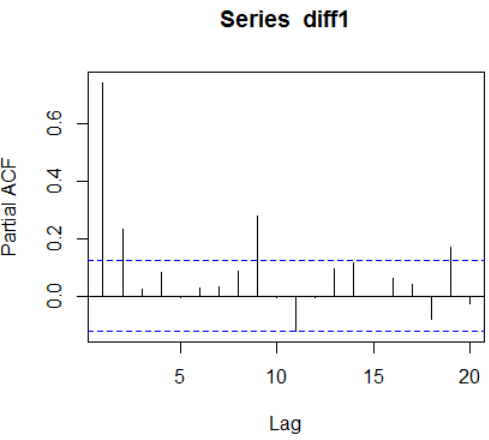
**Increasing lag.max=100:**



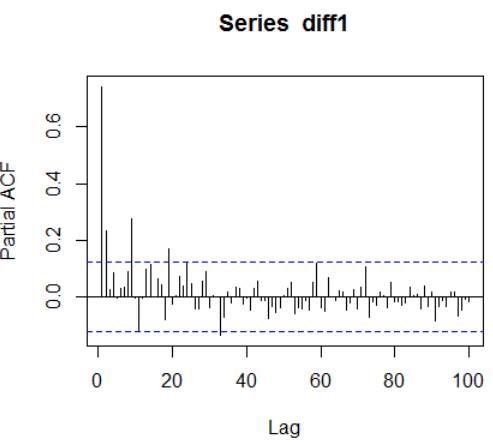
**We can see it’s decreasing. It’s hard to see how many lags.**

* 1. Compute the PACF values.





**Increasing lag.max=100:**

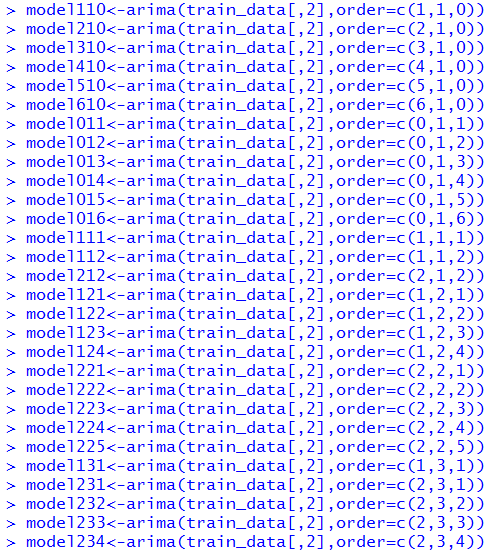


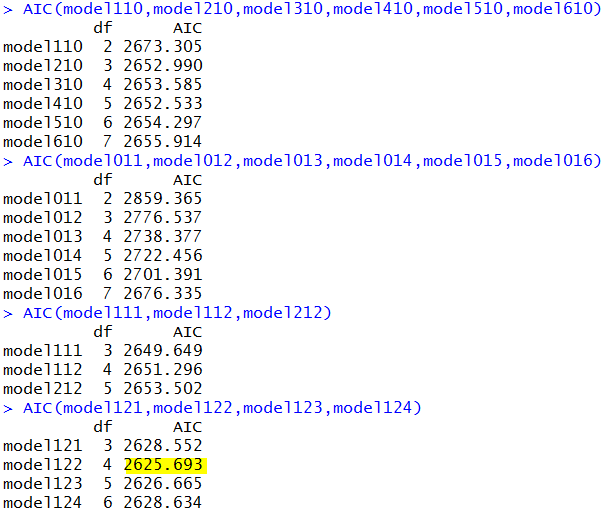
**It’s hard to see how many lags.**

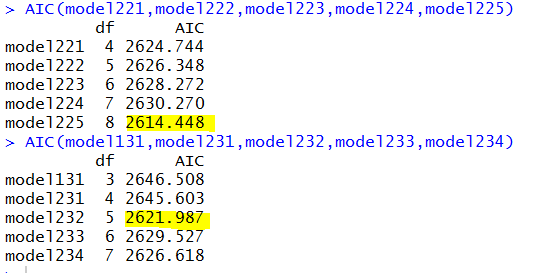
* 1. Determine the AR and MA lags in the data.

**From the ACF and PACF it’s hard to decide the lags for MA and AR. We just can see acf[2]=pacf[1]=0.744**

* 1. Suggest a ARIMA model(n,d,m) for this data.



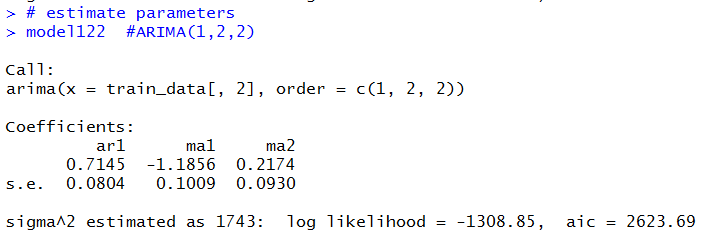




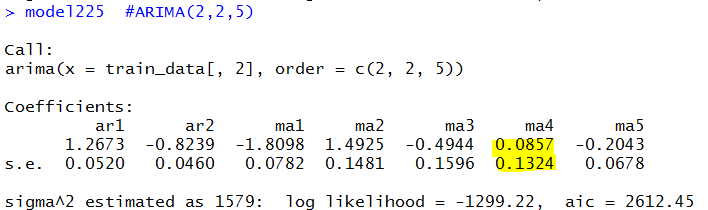
**I try these models, I found that ARIMA(1,2,2), ARIMA(2,2,5) and ARIMA(2,3,2) get the smallest AIC, which is 2625.693,2614.448 and 2621.987 separately.**

**So I suggest ARIMA(1,2,2), ARIMA(2,2,5) and ARIMA(2,3,2)**

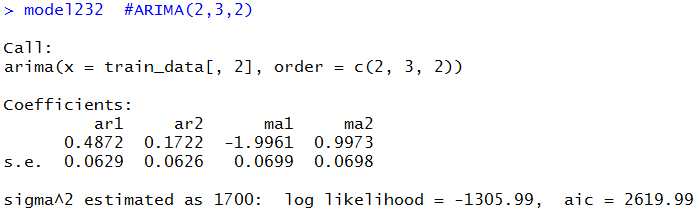
* 1. Estimate the AR and MA parameters for your ARIMA model.



**ar1=0.7145,ma1=-1.1856,ma2=0.2174**



**Model ARIMA(2,2,5), we can see ma4 the value 0.0857 contains 0 within 2 its standard error. 0.0857-0.2648<0, So I reject this model.**



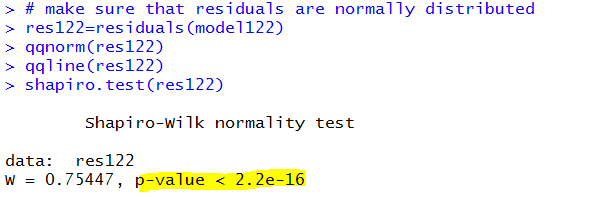
**ar1=0.4872,ar2=0.1722,ma1=-1.9961,ma2=0.9973**

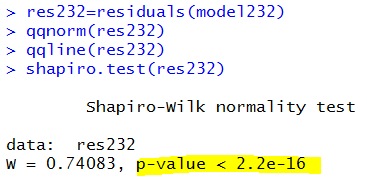
* 1. Compute the diagnostics data of your model.

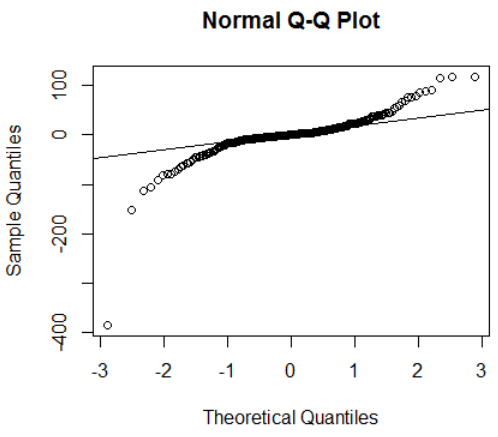
**From the upper result, we can get the residuals of those 3 suggested models as following:**

|  |  |  |
| --- | --- | --- |
|  | **LL** | **AIC** |
| **Model122:ARIMA(1,2,2)** | -1308.85 | 2623.69 |
| **Model232:ARIMA(2,3,2)** | -1305.99 | 2619.99 |

**From the result, I can’t tell which one is the best. So I will use both models.**





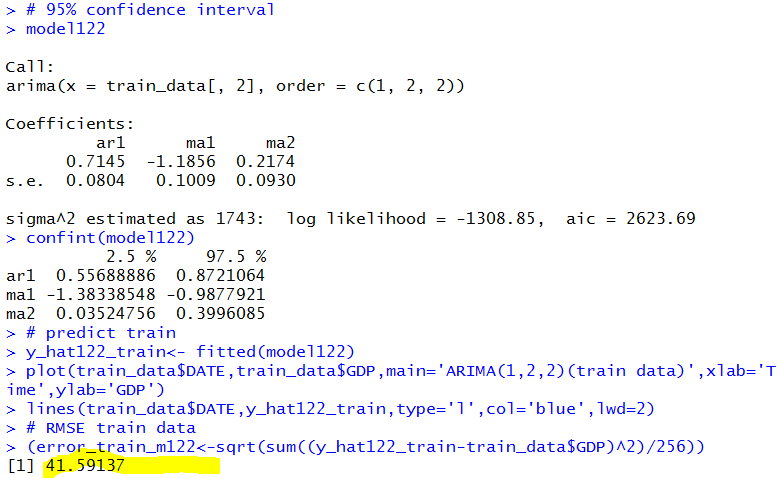


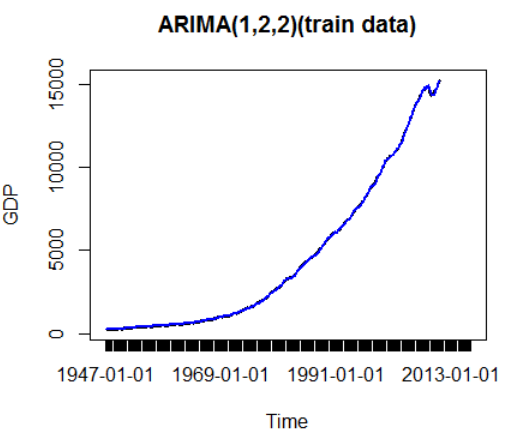
**Even though I found that all Q-Q Plot is not normal distributed, but p-value of shapiro.test are all significant. I assume their residuals are normal distributed.**

* 1. Compute the forecast and its RMSE.

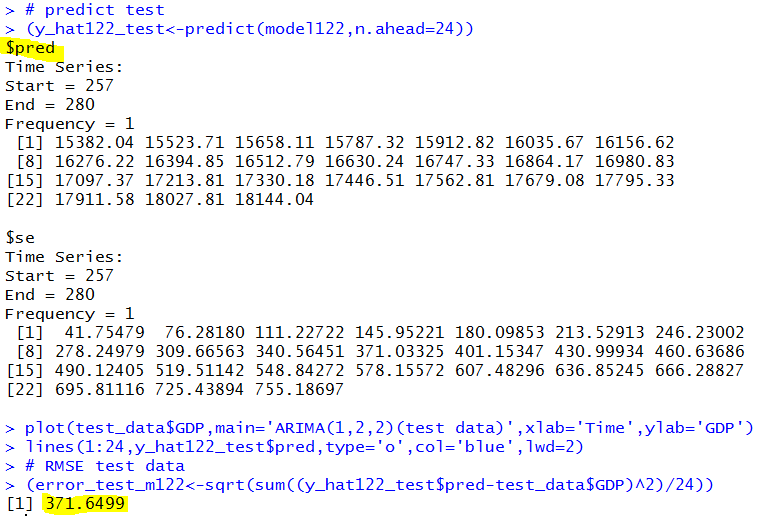
**ARIMA(1,2,2):**

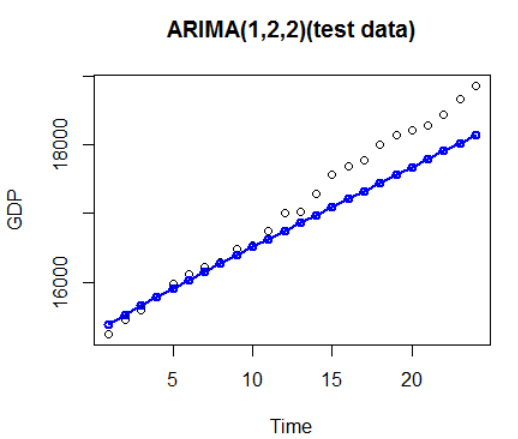
**i). Predict training data:**





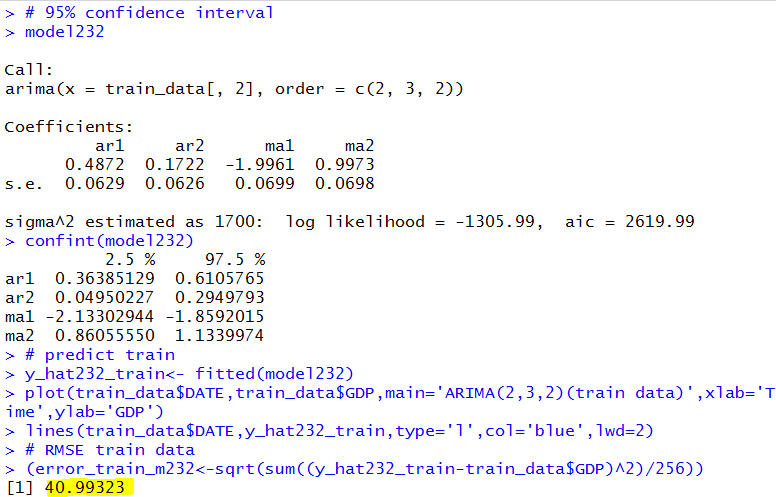
**ii). Predict test data**

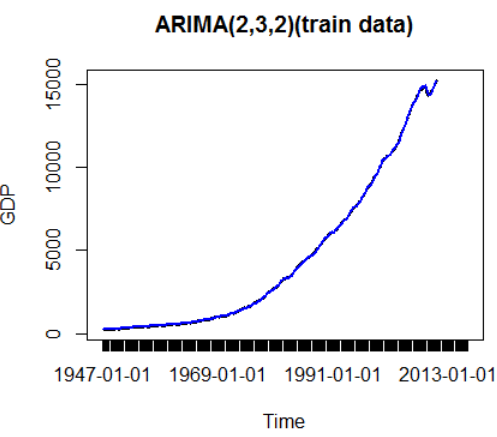




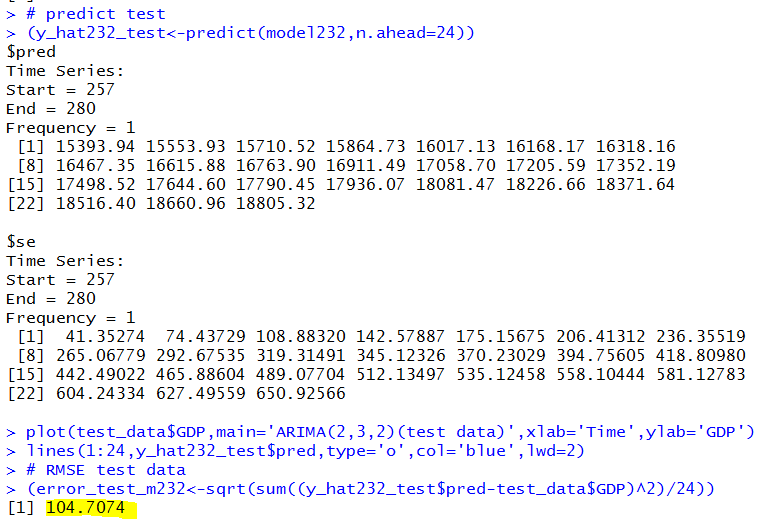
**ARIMA(2,3,2):**

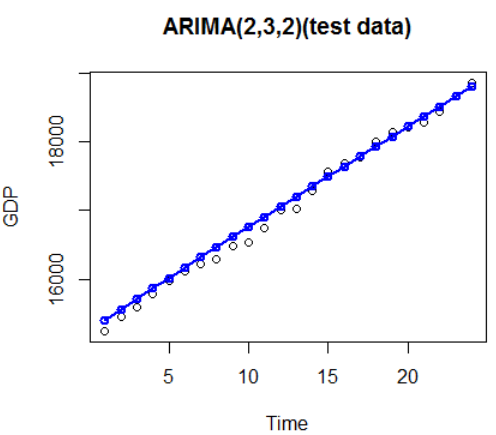
**i). Predict training data:**





**ii). Predict test data**





**From the RMSE, I can get that ARIMA(2,3,2) get a better result.**

**Summary:**

From the result, we can see that EWMA get the least RMSE for testing data, which is 94.46443. I would select EWMA for forecasting for this dataset. However, it’s very close with ARIMA. So ARIMA could always be the choice to consider of. We can use both EWMA and ARIMA to forecasting and the a more reliable conclusion.

Compare to all these five models, ARIMA and EWMA perform much better than Linear, quadratic, and exponential model, which is obvious that time series depends on the previous data, we can’t use the whole dataset to build regression model to forecast.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Time Series Model | Model Specification(y: GDP, x: Date(start:1947/01/01) | RMSE for the Testing Data |
| 1 | Linear | **y**=57.794\***x**-2918.300 | 4511.857 |
| 2 | Quadratic | **y**=0.348688\***x2**-31.818592\***x**+935.054657 | 490.5476 |
| 3 | Exponential | **y**=1.017814**x** \* 229.8038 | 9702.007 |
| 4 | EWMA | **y\_hat(t+1)** = 0.566 \* (diff\_y(t)-y\_hat(t)) + y\_hat(t) = 0.566\*diff\_y(t) + 0.434\*y\_hat(t) **diff\_y(t) =** y(t) - y(t-1) | 94.46443 |
| 5 | ARIMA | **arima(2,3,2)** | 104.7074 |