FORECASTING GDP OF USA USING ARIMA

Pendem Shiva Kumar

Master of Science in Data Science,
Indiana University
May' 24

Course: ECON-M 524(FINANCIAL ECONOMETRICS)

Instructor: Dr. Keli Xu

Introduction

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.

Objective (Research Question)

The project aims to forecast the GDP of the United States of America with dataset taken from the Federal Reserve Economic Data.

Theory

ARIMA, or autoregressive integrated moving average, is a statistical model used for time series forecasting. It is a popular and widely used method for forecasting a variety of economic and financial data, such as GDP, inflation, and stock prices.

The ARIMA model is a type of linear regression model that incorporates information about the past values of a time series in order to make more accurate predictions about its future values. The model is composed of three components: the autoregressive (AR) term, the differencing (I) term, and the moving average (MA) term. The AR term captures the autocorrelation in the data, the I term accounts for non-stationarity in the data, and the MA term models the random error in the data.

A nonseasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

- **p** is the number of autoregressive terms,
- **d** is the number of nonseasonal differences needed for stationarity, and
- **q** is the number of lagged forecast errors in the prediction equation.

The forecasting equation is constructed as follows. First, let y denote the d^{th} difference of Y, which means:

If d=0: $y_t = Y_t$

If d=1: $y_t = Y_t - Y_{t-1}$

If d=2: $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$

Note that the second difference of Y (the d=2 case) is not the difference from 2 periods ago. Rather, it is the *first-difference-of-the-first difference*, which is the discrete analog of a second derivative, i.e., the local acceleration of the series rather than its local trend.

In terms of y, the general forecasting equation is:

$\hat{y}_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$

Here the moving average parameters (θ's) are defined so that their signs are negative in the equation, following the convention introduced by Box and Jenkins. Some authors and software (including the R programming language) define them so that they have plus signs instead. When actual numbers are plugged into the equation, there is no ambiguity, but it's important to know which convention your software uses when you are reading the output. Often the parameters are denoted there by AR(1), AR(2), …, and MA(1), MA(2), … etc..

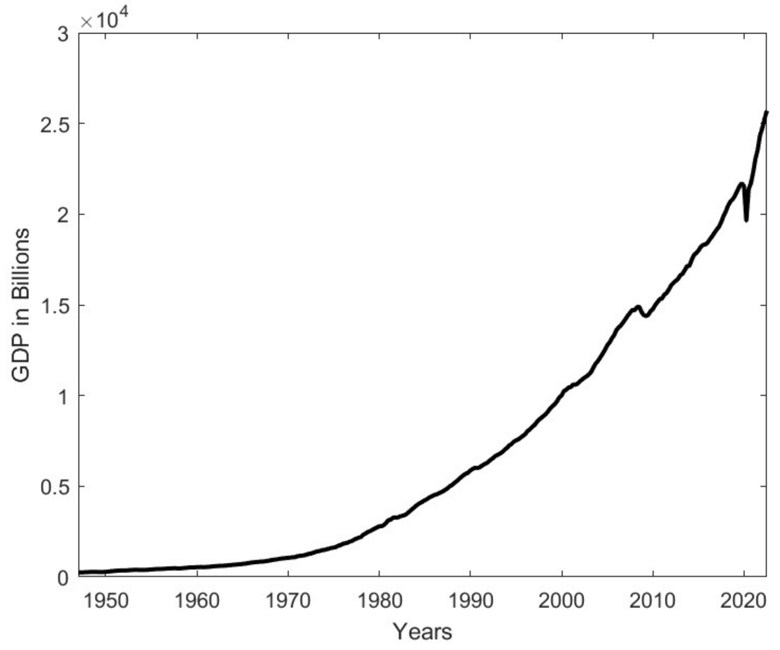
To identify the appropriate ARIMA model for Y, you begin by determining the order of differencing (d) needing to stationarize the series and remove the gross features of seasonality, perhaps in conjunction with a variance-stabilizing transformation such as logging or deflating. If you stop at this point and predict that the differenced series is constant, you have merely fitted a random walk or random trend model. However, the stationarized series may still have autocorrelated errors, suggesting that some number of AR terms ($p \geq 1$) and/or some number MA terms ($q \geq 1$) are also needed in the forecasting equation.

Data

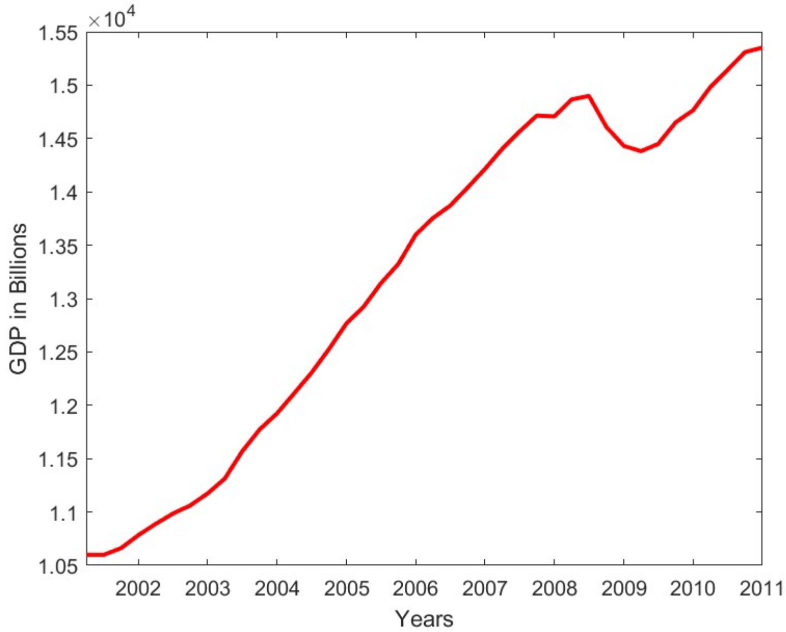
FRED provided the data in .csv form. The 'GDP.csv' contains 2 columns. 'DATE' and 'GDP'. The frequency of the GDP is quarterly i.e, GDP of every quarter from the beginning of January 1947 to 2nd Quarter of 2022.

	A	B
1	DATE	GDP
2	01-01-47	243.164
3	01-04-47	245.968
4	01-07-47	249.585
5	01-10-47	259.745
6	01-01-48	265.742
7	01-04-48	272.567
8	01-07-48	279.196
9	01-10-48	280.366
10	01-01-49	275.034
11	01-04-49	271.351
12	01-07-49	272.889

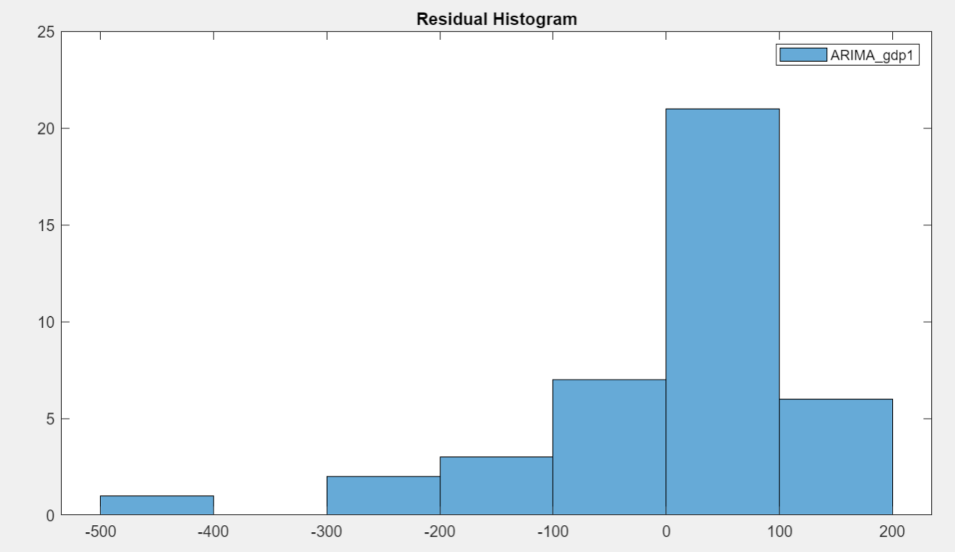
Research Method



The above figure displays the GDP of The United States from 1947 to 2nd Quarter of 2022. Note: The Y-axis of the figure is in Trillions. (1000 Billions).



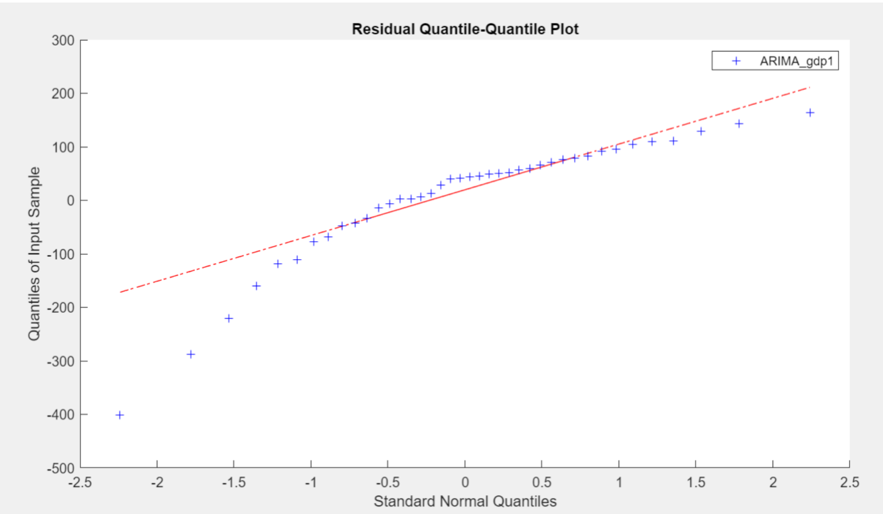
I am only considering data from year 1st Quarter of 2001 to the last Quarter of 2010. Using this decade of data to forecast the data from 2011 to 2022 and compare the forecasted series with the true values.



A residual histogram is a plot that shows the distribution of the residuals (the difference between the observed values and the predicted values) of a regression model. This plot can be useful for identifying any patterns or trends in the residuals, as well as assessing the overall fit of the model.

A well-fitting regression model should have residuals that are randomly distributed and centered around zero, with no discernible patterns or trends. If the residuals are not randomly distributed, this may indicate that the model is not accurately capturing the underlying relationships in the data.

Quantile-Quantile Plot



QQ plots, or quantile-quantile plots, are used to assess whether a dataset follows a specific theoretical distribution. This is important because many statistical methods assume that the data follows a particular distribution, such as a normal distribution. If the data does not follow the assumed distribution, the results of the statistical analysis may be inaccurate or misleading.

The residual Histogram plots and residual QQ plots are created in [Econometric Modeler](#) which is an application in MATLAB software primarily used for econometrics.

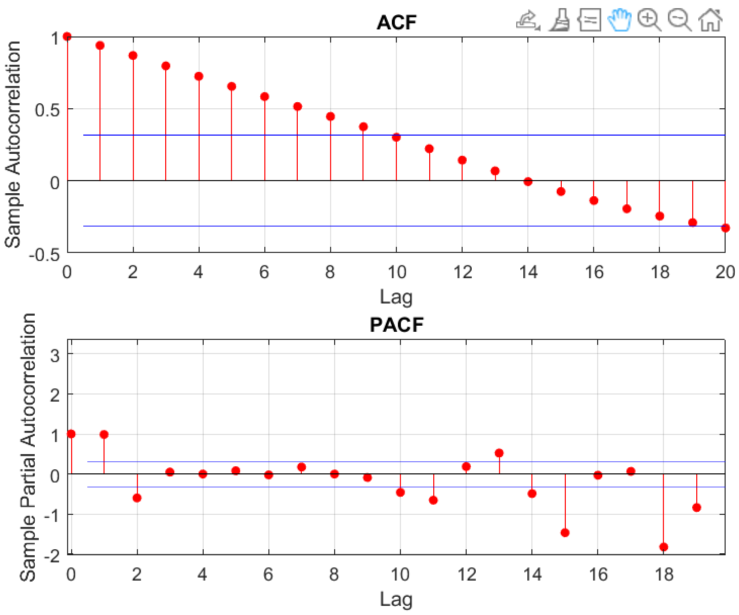
Performing ACF and PACF on the data (2001 to 2010).

ACF and PACF are acronyms that refer to two statistical tools used in time series analysis: the autocorrelation function (ACF) and the partial autocorrelation function (PACF).

The autocorrelation function (ACF) is a measure of the correlation between a time series and a lagged version of itself. It is used to identify the presence of autocorrelation, which is the tendency of a time series to be correlated with previous or future observations of itself. The ACF can be used to help identify the order of an autoregressive model, which is a type of statistical model that describes the time-varying behavior of a time series.

The partial autocorrelation function (PACF) is a measure of the partial correlation between a time series and a lagged version of itself, controlling for the values of the time series at all shorter lags. It is used to identify the order of a moving average model, which is a type of statistical model that describes the time-varying behavior of a time series by taking into account the past values of the error term.

Together, the ACF and PACF can be used to identify the appropriate order of an autoregressive-moving average (ARMA) model, which is a type of statistical model that combines both autoregressive and moving average components to capture the time-varying behavior of a time series.



The code below is used to create the ACF and PACF test.

```
figure
subplot(2,1,1);
autocorr(log(gdp));
title("ACF");
subplot(2,1,2);
```

```
parcorr(log(gdp));
title("PACF");
```

AIC and BIC Tests

AIC and BIC are two commonly used model selection criteria in statistics and machine learning. AIC stands for the Akaike information criterion, and BIC stands for the Bayesian information criterion.

Both AIC and BIC are used to compare the relative quality of different statistical models by taking into account their complexity and their ability to fit the data. These criteria are based on the principle that a good model should strike a balance between fitting the data well and being parsimonious, that is, having a relatively small number of parameters.

AIC and BIC are typically used in model selection to help identify the best-fitting model among a set of candidate models. They are used to compare the relative quality of different models by assigning a score to each model based on its fit to the data and its complexity. The model with the lowest AIC or BIC score is considered to be the best-fitting model.

In general, AIC and BIC are useful tools for model selection because they provide a way to compare the quality of different models in a consistent and objective manner. They can help researchers and analysts choose the best-fitting model for their data and avoid overfitting, which occurs when a model is overly complex and does not generalize well to new data.

There is no definitive answer to which model selection criterion is best between AIC and BIC. Both AIC and BIC are widely used model selection criteria that have their own advantages and disadvantages.

AIC is widely used because it has good theoretical properties and is easy to compute. However, AIC has some limitations, such as its reliance on large sample sizes and its inability to account for model uncertainty.

BIC has the advantage of being more conservative than AIC, which means it is less likely to overfit the data. However, BIC is more difficult to compute than AIC and can be sensitive to the choice of prior distribution.

By running this code I was able to generate the AIC and BIC tests for my data.

```
[aic,bic]= aicbic(LogLikelihood,2,40);
Model_table = "arima(0,1,0)";
aic_table = aic;
bic_table = bic;
m=2;
for i = 1:2
    for j = 1:2
        for k=1:2
            arima_model(m-1) = arima(i,j,k);
            [~,~,LoglikehoodE]            =            estimate(arima_model(m-
1),gdp,'display','off');
            [aic_inter,bic_inter] = aicbic(LoglikehoodE,2,250);
```

```
            Model_table(m)                                                              =
['arima(',num2str(i),',',num2str(j),',',num2str(k),')'];
            aic_table(m) = aic_inter;
            bic_table(m) = bic_inter;
            m=m+1;
        end
    end
end
Comparision = table(Model_table',aic_table',bic_table');
disp(Comparision);
```

Below are the AIC and BIC tests:

Var1	Var2	Var3
"arima(0,1,0)"	497.18	500.56
"arima(1,1,1)"	471.74	478.78
"arima(1,1,2)"	471.68	478.73
"arima(1,2,1)"	474.73	481.77
"arima(1,2,2)"	474.3	481.34
"arima(2,1,1)"	468.96	476.01
"arima(2,1,2)"	468	475.04
"arima(2,2,1)"	474.51	481.56
"arima(2,2,2)"	474.33	481.37

We need to find the lowest AIC or BIC values and go ahead. One can eyeball the min value but it is impractical for large simulations. I've written a code which gives the minimum AIC or BIC values.

Below is minimum AIC value index and corresponding model:

```
AICMin=aic_table(1);
AICindex = 1;
for i=1:length(aic_table)
    if(aic_table(i)<AICMin)
        AICindex=i;
        AICMin=aic_table(i);
    end
end

disp(aic_table);
```

```
  497.1784   471.7383   471.6850   474.7305   474.3005   468.9639   467.9987   474.5128   474.3291
```

```
disp("Lowest AIC value at:");
```

```
Lowest AIC value at:
```

```
disp(AICindex);
```

```
     7
```

```
disp(Model_table(AICindex));
```

```
arima(2,1,2)
```

Below is minimum BIC value index and corresponding model:

```
BICMin=bic_table(1);
BICindex = 1;
for i=1:length(bic_table)
    if(bic_table(i)<BICMin)
        BICindex=i;
        BICMin=bic_table(i);
    end
end
|
disp(bic_table);
```

```
  500.5562   478.7812   478.7279   481.7735   481.3434   476.0068   475.0417   481.5557   481.3720
```

```
disp("Lowest BIC value at:");
```

```
Lowest BIC value at:
```

```
disp(BICindex);
```

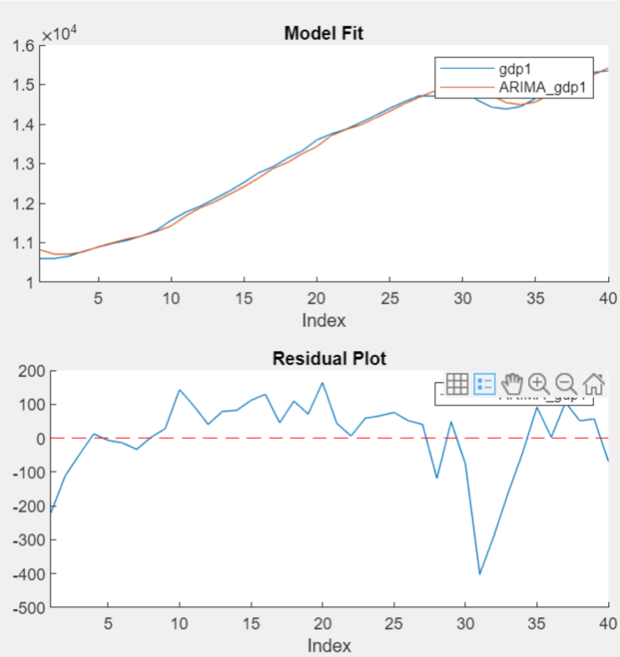
```
     7
```

```
disp(Model_table(BICindex));
```

```
arima(2,1,2)
```

In our case both AIC and BIC corresponding models are same i.e., arima(2,1,2).

One can also use Econometric Modeler application in MATLAB in order to do AIC and BIC tests. Below is the arima(0,1,0) fit to the data. One can try to go through all variations of arima model and do the analysis.



Parameters

Parameter	Value	Standard Error	t Statistic	P-Value
Constant	110.5219	20.8881	5.2911	1.2155e-07
Variance	1.2505e+04	2.4188e+03	5.1700	2.3413e-07

Goodness of Fit

Measure	Value
AIC	497.1784
BIC	500.5056

Implementing Theoretical ARIMA model:

Theoretical ARIMAor (Random Walk) model is arima(0,1,0) .

An ARIMA(0, 1, 0) series, when differenced once, becomes an ARMA(0, 0), which is random, uncorrelated, noise. If $X_1, X_2, X_3, \ldots$ are the random variables in the series, this means that

$$X_{i+1} - X_i = \epsilon_{i+1}$$

where $\epsilon_1, \epsilon_2, \ldots$ is a sequence of centered, uncorrelated random variables. Rearranging

$$X_{i+1} = X_i + \epsilon_i$$

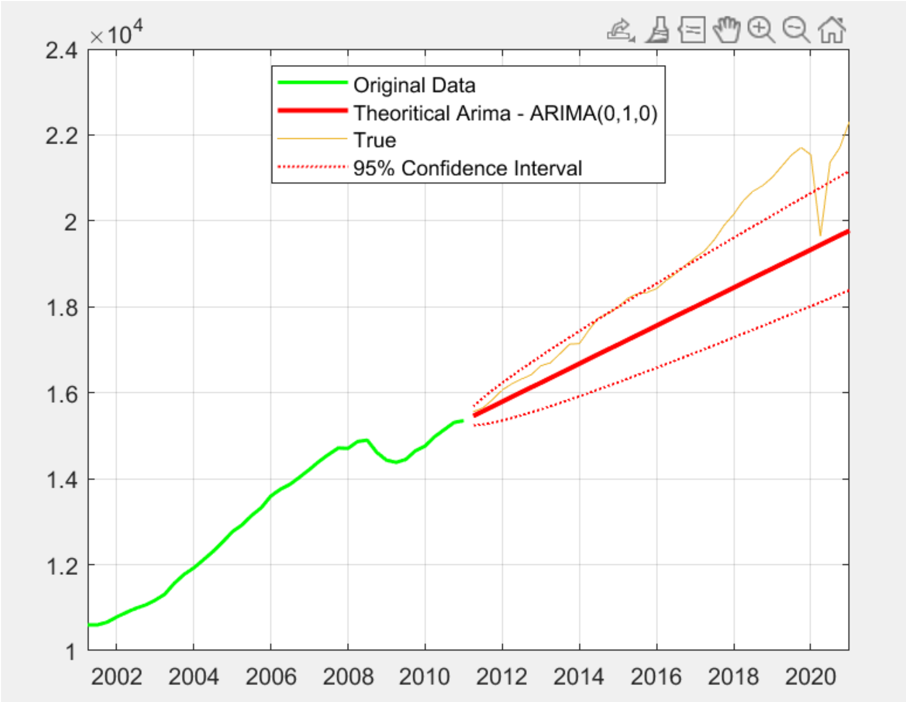
reveals that we have a random walk.

In matlab this is the command I've used:

```
ARIMA_Theo= arima(0,1,0); %Theoretical ARIMA model
[ARIMA_Theo1,~,LogLikelihood]= estimate(ARIMA_Theo, gdp);
```

Inorder to plot the the forecast of the random walk ARIMA and compare with the original data.

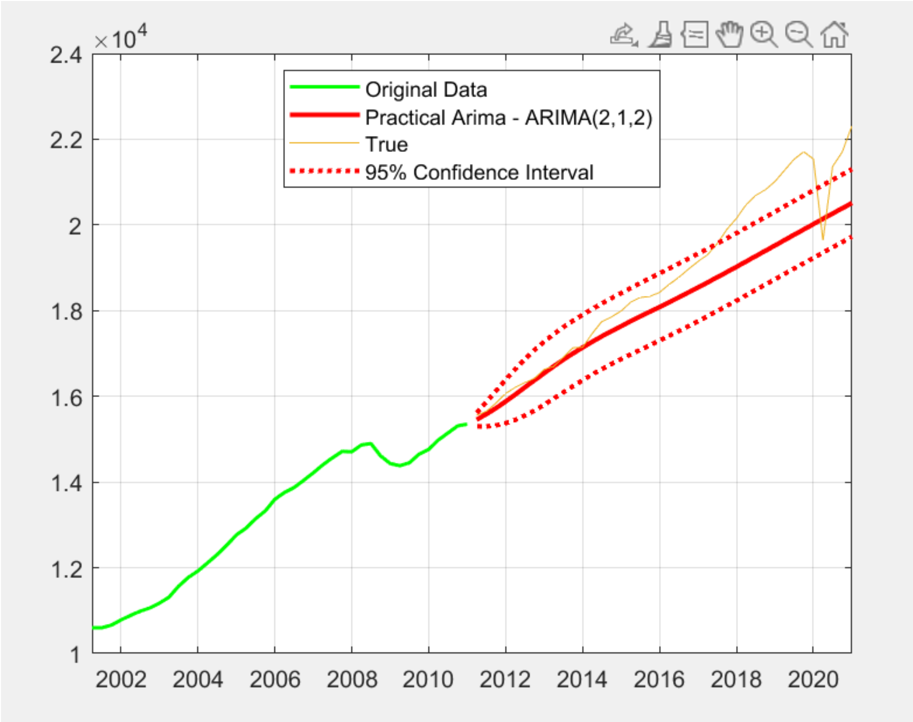
```
newlength=40;
%
% % Now, let us plot the forecasted data for ARIMA(0,1,0) model along with a
% % 95 percent confidence interval.


[Final_Theo_Forecast,ymse1]= forecast(ARIMA_Theo1, newlength,'Y0', gdp);
lower_theo= Final_Theo_Forecast - 1.96*sqrt(ymse1);
upper_theo= Final_Theo_Forecast + 1.96*sqrt(ymse1);
figure
h1 = plot(date,gdp,Color='g', LineWidth=1.5);
hold on
h2    =    plot(all_date(258:257+newlength),Final_Theo_Forecast,'LineWidth',2,
Color='r');
h3 = plot(all_date(258:257+newlength),all_gdp(258:257+newlength));
h4 = plot(all_date(258:257+newlength),lower_theo,'r:','LineWidth',1);
h5 = plot(all_date(258:257+newlength),upper_theo,'r:','LineWidth',1);
legend([h1,h2,h3,h4],"Original Data ","Theoritical Arima - ARIMA(0,1,0)","True
","95% Confidence Interval","Location","Best");
grid on
hold off
```



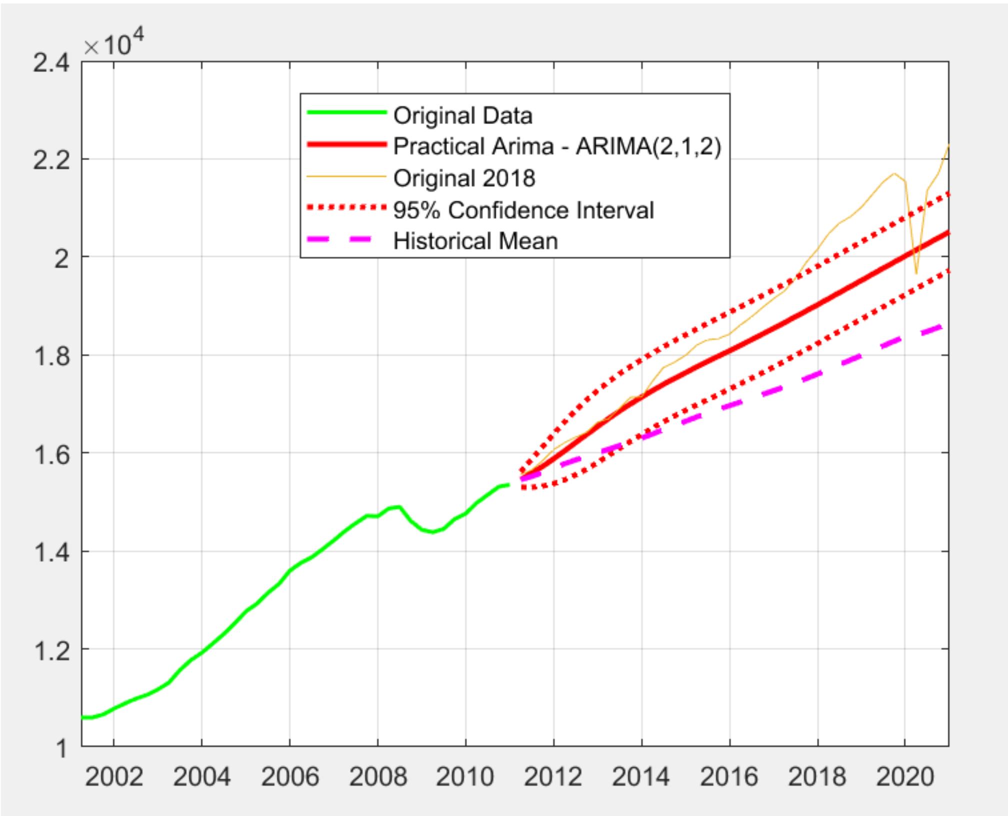
Corresponding plot obtained from the code.

We will now plot the Practical ARIMA model i.e., the best ARIMA model for our data. Below is the code I've used to generate the plot of arima(2,1,2).

```
[Final_Prac_Forecast,ymse1]= forecast(ARIMA_Prac1, newlength,'Y0', gdp);
lower_prac= Final_Prac_Forecast - 1.96*sqrt(ymse1);
upper_prac= Final_Prac_Forecast + 1.96*sqrt(ymse1);
figure
h1 = plot(date,gdp,Color='g', LineWidth=1.5);
hold on
h2    =    plot(all_date(258:257+newlength),Final_Prac_Forecast,'LineWidth',2,
Color='r');
h3 = plot(all_date(258:257+newlength),all_gdp(258:257+newlength));
h4 = plot(all_date(258:257+newlength),lower_prac,'r:','LineWidth',2);
h5 = plot(all_date(258:257+newlength),upper_prac,'r:','LineWidth',2);
legend([h1,h2,h3,h4],"Original Data","Practical Arima - ARIMA(2,1,2)","Original
2018","95% Confidence Interval","Location","Best");
grid on
hold off
```



Comparing the plot with the historical mean of the data.



How do I know that my arima(2,1,2) model did better than arima(0,1,0) *(Random walk)*?

I've computed a parameter called root mean square error or root mean square deviation (RMSD)

RMSD is defined as

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \hat{x}_i)^2}{N}}$$

RMSD = root-mean-square deviation
i = variable i
N = number of non-missing data points
x_i = actual observations time series
$\hat{x}_i$ = estimated time series

infer(predicted, actual) generates the residuals

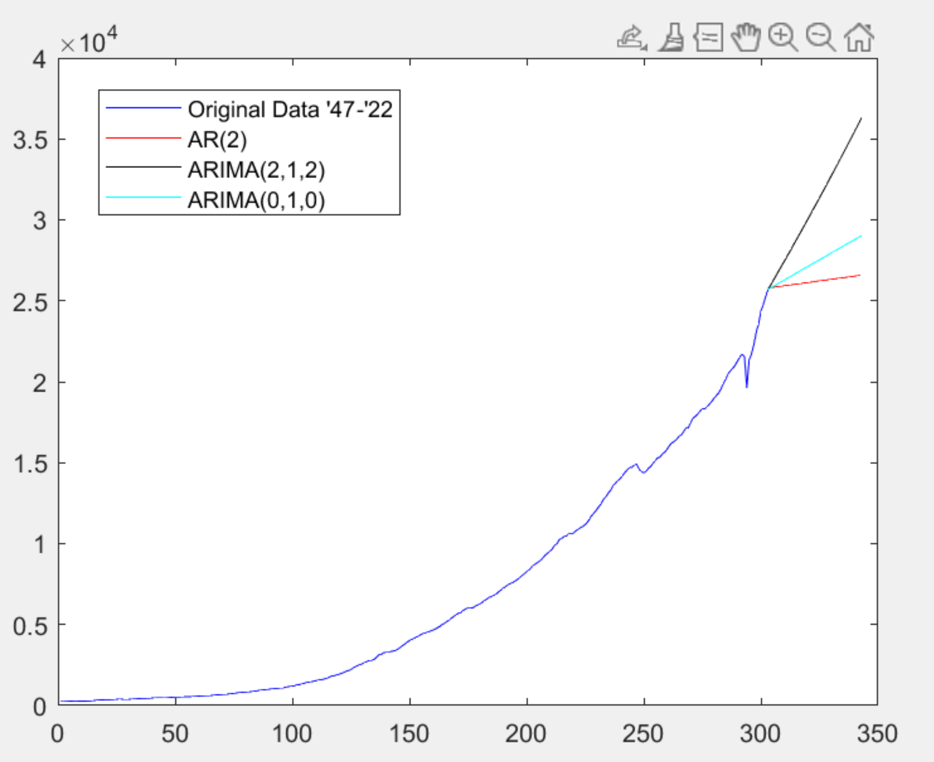
Here, Prac1 is the arima(2,1,2) and Theo1 is arima(0,1,0)

```
residual= infer(ARIMA_Theo1, gdp);
root_mean_square_error_Theo1 = sqrt(sum(residual.^2))/length(residual)
residual2 = infer(ARIMA_Prac1, gdp );
root_mean_square_error_Prac1 = sqrt(sum(residual2.^2))/length(residual2)
```

```
root_mean_square_error_Theo1 =

   18.1842


root_mean_square_error_Prac1 =

   12.6367
```

We can clearly observe that our model does better than a random walk by comparing the values of root mean squared error.

Forecasting GDP of next 10 years using ARIMA(2,1,2).



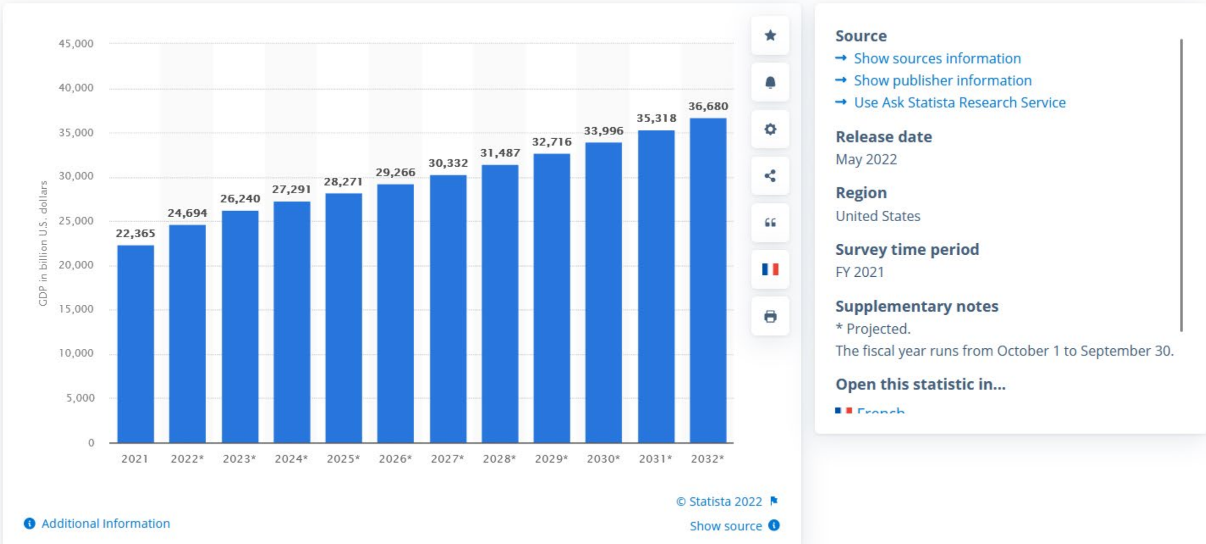
There are no true values for me to compare inorder to know that my model is performing good or not. I've compared my results with the other forecasts.

I've compared my forecast with Statista company's forecast of the GDP of US for the next 10 years.

Economy & Politics › Economy

Forecast of the U.S. Gross Domestic Product (GDP) for fiscal years 2021 to 2032

(in billion U.S. dollars)



Our models forecasting is similar to Statista's forecast. One can't simply rely on this arima(2,1,2) model to forecast GDP forever.

GDP forecasting by economists is done by first collect data on various economic indicators, such as employment levels, consumer spending, and business investment. They then use statistical methods and economic models to analyze this data and make predictions about how these indicators will change in the future. For example, they might use regression analysis to identify trends in the data and make projections based on those trends. Once they have created a model, they can use it to estimate the future growth of the economy and report the results as a GDP forecast.

Conclusion.

I have used ARIMA Model in my project because it is relatively simple to understand and implement. It is also a flexible model that can be used to analyze a wide range of time series data. In addition, the ARIMA model can help to identify and remove any trends or seasonal patterns in the data, which can improve the accuracy of the forecasts. Another advantage is that the ARIMA model can incorporate exogenous variables, which are external factors that may influence the time series data. This allows the model to take into account factors such as economic policy or natural disasters that may affect the data. Although my model does a good job of forecasting from year 2011 to 2022. **One couldn't simply rely on this model** for forecasting the GDP forever.

Computation of macroeconomic factor such as GDP is always hard. Researchers might use dynamic stochastic general equilibrium (DSGE) models, which are a type of macroeconomic model that is often used to study the interactions between different sectors of the economy. These models can help to capture the complex relationships between various economic variables, such as interest rates, inflation, and employment, and can provide more accurate forecasts of GDP growth.

One of the main limitations of GDP forecasting is that it is inherently uncertain and subject to error. Economic data is often revised after it is initially reported, which can cause forecast errors. In addition, economic models are based on assumptions about the future, and these assumptions may not always hold true. For example, a forecast may not account for unexpected events, such as natural disasters or political instability, which can have a significant impact on the economy. Furthermore, the accuracy of a forecast can also be affected by the quality and availability of data, as well as the reliability of the underlying statistical methods and economic models.

Appendix

Code:

```
data = readtable("GDP.csv")
all_gdp = data.GDP
all_date = data.DATE
%historcial mean calculation
hm = zeros(303,1)
hm = all_gdp
for i=1:46
hm(i+257) = mean(all_gdp(257:257+i));
end
MSE_hm = mean((hm(258:257+newlength)-all_gdp(258:257+newlength)).^2)

%testing stationary
% Y = log(gdp)
% h = adftest(Y,Model="TS",Lags=0:2)


%values from 2001 to now
gdp = data.GDP(218:257)
date = data.DATE(218:257)
plot(date,gdp, 'r', LineWidth=2);
xlabel('Years')
ylabel('GDP in Billions')


figure
subplot(2,1,1);
autocorr(log(gdp));
title("ACF");
subplot(2,1,2);
parcorr(log(gdp));
title("PACF");

[aic,bic]= aicbic(LogLikelihood,2,40);
Model_table = "arima(0,1,0)";
aic_table = aic;
bic_table = bic;

%computation of AIC and BIC tests
m=2;
for i = 1:2
    for j = 1:2
        for k=1:2
            arima_model(m-1) = arima(i,j,k);
            [~,~,LoglikehoodE] = estimate(arima_model(m-1),gdp,'display','off');
            [aic_inter,bic_inter] = aicbic(LoglikehoodE,2,250);
            Model_table(m) =
['arima(',num2str(i),',',num2str(j),',',num2str(k),')'];
            aic_table(m) = aic_inter;
            bic_table(m) = bic_inter;
```

```
            m=m+1;
        end
    end
end
Comparision = table(Model_table',aic_table',bic_table');
disp(Comparision);
%AIC
AICMin=aic_table(1);
AICindex = 1;
for i=1:length(aic_table)
    if(aic_table(i)<AICMin)
        AICindex=i;
        AICMin=aic_table(i);
    end
end
disp(aic_table);

disp("Lowest AIC value at:");
disp(AICindex);
disp(Model_table(AICindex));

%BIC
BICMin=bic_table(1);
BICindex = 1;
for i=1:length(bic_table)
    if(bic_table(i)<BICMin)
        BICindex=i;
        BICMin=bic_table(i);
    end
end
disp(bic_table);
disp("Lowest BIC value at:");
disp(BICindex);
disp(Model_table(BICindex));


ARIMA_Theo= arima(0,1,0); %Theoretical ARIMA model
[ARIMA_Theo1,~,LogLikelihood]= estimate(ARIMA_Theo, gdp);

ARIMA_Prac= arima(2,1,2);%Practical ARIMA model obtained after the AIC BIC tests
[ARIMA_Prac1,~,LogLikelihood2]= estimate(ARIMA_Prac, gdp);

%RMS values
residual= infer(ARIMA_Theo1, gdp);     % This line basically generate the
difference in the values b/w the actual and fitted model.
root_mean_square_error_Theo1 = sqrt(sum(residual.^2))/length(residual)
residual2 = infer(ARIMA_Prac1, gdp );
root_mean_square_error_Prac1 = sqrt(sum(residual2.^2))/length(residual2)
prediction= gdp +residual;
prediction2 = gdp + residual2;




%FORECASTING THE FUTURE:
%IN THIS SECTION, WE WILL BE FORECASTING THE FUTURE DATA FOR THREE MONTHS AND
%COMPARE IT WITH THE ORIGINAL VALUES TO GET A BETTER UNDERSTANDING.
```

```
newlength=40;
%
% % Now, let us plot the forecasted data for ARIMA(0,1,0) model along with a
% % 95 percent confidence interval.

[Final_Theo_Forecast,ymse1]= forecast(ARIMA_Theo1, newlength,'Y0', gdp);
lower_theo= Final_Theo_Forecast - 1.96*sqrt(ymse1);
upper_theo= Final_Theo_Forecast + 1.96*sqrt(ymse1);
figure
h1 = plot(date,gdp,Color='g', LineWidth=1.5);
hold on
h2 = plot(all_date(258:257+newlength),Final_Theo_Forecast,'LineWidth',2,
Color='r');
h3 = plot(all_date(258:257+newlength),all_gdp(258:257+newlength));
h4 = plot(all_date(258:257+newlength),lower_theo,'r:','LineWidth',1);
h5 = plot(all_date(258:257+newlength),upper_theo,'r:','LineWidth',1);
legend([h1,h2,h3,h4],"Original Data ","Theoritical Arima - ARIMA(0,1,0)","True
","95% Confidence Interval","Location","Best");
grid on
hold off

%Practical ARIMA
[Final_Prac_Forecast,ymse1]= forecast(ARIMA_Prac1, newlength,'Y0', gdp);
lower_prac= Final_Prac_Forecast - 1.96*sqrt(ymse1);
upper_prac= Final_Prac_Forecast + 1.96*sqrt(ymse1);
figure
h1 = plot(date,gdp,Color='g', LineWidth=1.5);
hold on
h2 = plot(all_date(258:257+newlength),Final_Prac_Forecast,'LineWidth',2,
Color='r');
h3 = plot(all_date(258:257+newlength),all_gdp(258:257+newlength));
h4 = plot(all_date(258:257+newlength),lower_prac,'r:','LineWidth',2);
h5 = plot(all_date(258:257+newlength),upper_prac,'r:','LineWidth',2);
h6 = plot(all_date(258:257+newlength),hm(258:257+newlength),'LineWidth',2,
Color='m', LineStyle='--');
legend([h1,h2,h3,h4, h6],"Original Data","Practical Arima -
ARIMA(2,1,2)","Original 2018","95% Confidence Interval", "Historical
Mean","Location","Best");
grid on
hold off


%Historical Mean

figure
hold on
k1 = plot(all_date,all_gdp,'LineWidth',2, Color='r');
k2 =  plot(all_date(258:257+newlength),hm(258:257+newlength),'LineWidth',2,
Color='m', LineStyle='--');
legend([k1,k2],"Original Data ","Historical Mean", "Location","Best");
grid on
hold off


%%% Additional forecasting for 10 years
%AR
sys_2 = arima(2,0,0);
Md1_AR = estimate(sys_2,all_gdp);
yf = forecast(Md1_AR,40,'Y0',all_gdp); % Forecast 200 points ahead
```

```
plot(1:length(all_gdp),all_gdp,'b',length(all_gdp):length(all_gdp)+length(yf),[all
_gdp(end);yf],'r')

hold on
%MA
% sys_1 = arima(0,0,2);
% Md1_MA = estimate(sys_1,all_gdp);
% yf = forecast(Md1_MA,40,'Y0',all_gdp); % Forecast 200 points ahead
% plot(length(all_gdp):length(all_gdp)+length(yf),[all_gdp(end);yf],'m')

%ARMA
% sys_2 = arima(2,0,2);
% Md1_ARMA = estimate(sys_2,all_gdp)
% yf = forecast(Md1_ARMA,40,'Y0',all_gdp); % Forecast 200 points ahead
%
plot(1:length(all_gdp),all_gdp,'b',length(all_gdp):length(all_gdp)+length(yf),[all
_gdp(end);yf],'g'), legend('measured','forecasted')

%ARIMA
sys_3 = arima(2,1,2);
Md1_ARIMA = estimate(sys_3,all_gdp);
yf = forecast(Md1_ARIMA,40,'Y0',all_gdp);
 plot(length(all_gdp):length(all_gdp)+length(yf),[all_gdp(end);yf],'k')

sys_3 = arima(0,1,0);
Md1_ARIMA = estimate(sys_3,all_gdp);
yf = forecast(Md1_ARIMA,40,'Y0',all_gdp);
 plot(length(all_gdp):length(all_gdp)+length(yf),[all_gdp(end);yf],'c')
% legend([h1,h2,h3],"ARIMA(2,0,0)","ARIMA(2,1,2)","ARIMA(0,1,0)");
hold off
```

References:

https://www.investopedia.com/terms/g/gdp.asp

https://fred.stlouisfed.org/series/GDP

https://people.duke.edu/~rnau/411arim.htm#pdq

https://stats.stackexchange.com/questions/577/is-there-any-reason-to-prefer-the-aic-or-bic-over-the-other

https://en.wikipedia.org/wiki/Dynamic_stochastic_general_equilibrium

https://ca.indeed.com/career-advice/career-development/economic-forecasting

https://www.investopedia.com/terms/e/economic-forecasting.asp

https://www.statista.com/statistics/216985/forecast-of-us-gross-domestic-product/