ECON 427: Homework #3

Chart, line chart, scatter chart

Description automatically generated

Figure 1.1

1. There are seasonal fluctuations observed over the years except for around years 2008-09 and after 2010. Also, the trend of the time series is shown going in an upward direction. The GDP is observed to be the typically highest in the Q3 of each year and lowest in the Q1 of each year.
2. ACF

Chart

Description automatically generated

Figure 3.1

PACF (Partial Autocorrelation)

Chart

Description automatically generated

Figure 3.2

The ACF graph of the raw data shows that there is a strong autocorrelation present in the data as more than 5% of the data is outside the dotted lines. Whereas, PACF shows the autocorrelation at an individual lag, controlling for previous data/lags. In the PACF graph, there is some sort of seasonal variation visible at lags 1,2, and 3 which got reduced drastically at lag 4. Based on the model (PACF), we should use 2 lags to fit an AR(*p*) model to the data as it is showing autocorrelation without taking into consideration previous lags.

|  |  |
| --- | --- |
| **Model** | **AIC** |
| ARIMA(1,0,0) | 935.29 |
| ARIMA(2,0,0) | 930.38 |
| ARIMA(3,0,0) | 930.28 |
| ARIMA(4,0,0) | 909.76 |
| **ARIMA(5,0,0)** | **872.48** |
| ARIMA(6,0,0) | 872.57 |

Table 4.1

In the AR(*p*) model, *p* determines the number/order of lags.The best AR(*p*) model for the data is ARIMA(5,0,0) as it minimizes the ‘aic’ to the maximum among all other AR(*p*) models.

The steps that I took to manually determine the best AR(*p*) model are as follows:-

* At first, I created 6 different values using the ‘arima’ function where I changed the number of lags (*p*) of each model without making changes in its MA (*q*) and Seasonal (*d*) components.
* Then I ran all 6 models and compared its ‘aic’ to figure out the model that minimizes the ‘aic’.
* As observed in the Table 4.1, ARIMA(5,0,0) has the lowest AIC.
* The coefficients and other variables of the best model are as follows:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Intercept | ar1 | ar2 | ar3 | ar4 | ar5 |
| Coefficients | 3969.5091 | 0.8623 | -0.0702 | -0.0202 | 0.9275 | -0.7096 |
| Standard Error | 446.0753 | 0.0879 | 0.0784 | 0.0834 | 0.0802 | 0.0876 |
| AIC | 872.48 | | | | | |

Table 4.2

Chart

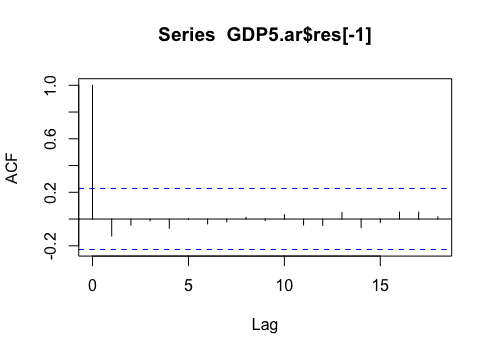
Description automatically generated Chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated Chart

Description automatically generated

 Chart

Description automatically generated

For this question, I exported residuals of all models to see the differences and in models with *p* < 5, some number of lags are outside the dotted lines which tells us that those models are strongly autocorrelated. Whereas the model with *p* = 5 looks like an appropriate correlogram of the residual showing it as the white noise. Also, the model with *p* = 6 is not much different than *p* = 5 and the difference in their aic is only 0.09. there doesn’t seem to be a seasonal variation in the model with *p* = 5.

|  |  |
| --- | --- |
| **Model** | **AIC** |
| ARIMA(0,0,1) | 1059.3 |
| ARIMA(0,0,2) | 985.96 |
| ARIMA(0,0,3) | 973.06 |
| ARIMA(0,0,4) | 964.57 |
| ARIMA(0,0,5) | 949 |
| ARIMA(0,0,6) | 927.93 |
| ARIMA(0,0,7) | 927.11 |
| ARIMA(0,0,8) | 919.9 |
| ARIMA(0,0,9) | 921.02 |
| **ARIMA(0,0,10)** | **907.89** |
| ARIMA(0,0,11) | 909.87 |
| ARIMA(0,0,12) | 907.95 |

Table 6.1

In the MA(*q*) model, *q* determines the number/order of lags.The best MA(*q*) model for the data is ARIMA(0,0,10) as it minimizes the ‘aic’ to the maximum among all other MA(*q*) models.

The steps that I took to manually determine the best MA(*q*) model are as follows:-

* At first, I created 12 different values using the ‘arima’ function where I changed the number of lags (*q*) of each model without making changes in its AR(*p*) and Seasonal (*d*) components.
* Then I ran all 12 models and compared its ‘aic’ to figure out the model which has minimum ‘aic’.
* As observed in the Table 6.1, ARIMA(0,0,10) has the lowest AIC.
* The coefficients and other variables of the best model are as follows:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Intercept | ma1 | ma2 | ma3 | ma4 | ma5 | ma6 | ma7 | ma8 | ma9 | ma10 |
| Coefficients | 4046.3991 | 1.0915 | 1.2559 | 0.8819 | 1.7501 | 1.8666 | 1.6872 | 0.6780 | 0.9325 | 0.8080 | 0.8809 |
| Standard Error | 105.9656 | 0.1964 | 0.5754 | 0.3230 | 0.3227 | 0.5292 | 0.8599 | 0.3479 | 0.2841 | 0.3302 | 0.4742 |
| AIC | 907.89 | | | | | | | | | | |

Table 6.2

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

For this question, I exported residuals of all models to see the differences and in models with *q* < 10, some number of lags are outside the dotted lines which tells us that those models are strongly autocorrelated. Whereas the model with *q* = 10 looks like an appropriate correlogram of the residual showing it as the white noise. Also, the model with *q* = 11 and *q* = 12 are almost similar to than *q* = 10. Along with that, residual of the model with *q* = 10 has some seasonal variation at lags 5,10,15...

1. The coefficients and other variables of the model are as follows:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Best Model | ARIMA(5,0,0) with non-zero mean | | | | | |
|  | Intercept | ar1 | ar2 | ar3 | ar4 | ar5 |
| Coefficients | 3969.5091 | 0.8623 | -0.0702 | -0.0202 | 0.9275 | -0.7096 |
| Standard Error | 446.0753 | 0.0879 | 0.0784 | 0.0834 | 0.0802 | 0.0876 |
| AIC | 872.48 | | | | | |

Table 8

Chart

Description automatically generated

Figure 9

Figure 9 displays the correlogram of the residuals for the best version of . The best version of is which means that in the model, AR=5, MA=0, and Seasonal = FALSE. The residual of the model looks like white noise as all the lags are within the dotted lines (confidence interval) and there is no autocorrelation visible at all. Also, there is no specific seasonal variation visible in the diagram.

1. The coefficients and other variables of the model are as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Best Model | ARIMA(4,1,0) with drift | | | |
|  | ar1 | ar2 | ar3 | ar4 |
| Coefficients | -0.1388 | -0.2094 | -0.2308 | 0.7004 |
| Standard Errors | 0.0884 | 0.0783 | 0.0963 | 0.0861 |
| AIC | 853.31 | | | |

Table 10

Chart

Description automatically generated

Figure 11

Figure 11 displays the correlogram of the residuals for the best version of . The best version of is which means that non-seasonal part of the model has AR=4, MA=1, and Seasonal = 0. The residual of the model looks like white noise as all the lags are within the dotted lines (confidence interval) and there is no autocorrelation visible at all. Also, there is no specific seasonal variation visible in the diagram.

1. Seasonally differenced series one-time

Chart

Description automatically generated

Figure 12

1. The coefficients and other variables of the ARIMA(1,0,0)(2,1,0)[4] model are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| Best Model | ARIMA(1,0,0)(2,1,0)[4] with drift | | |
|  | ar1 | sar1 | sar2 |
| Coefficients | 0.8304 | -0.2149 | 0.0359 |
| Standard Errors | 0.0961 | 0.2880 | 0.2394 |
| AIC | 815.32 | | |

Table 13

1. In the model, *p* determines the order of lags for Autoregressive models in the non-seasonal component, *q* determines the order of lags for Moving Average models in the non-seasonal component, and *d* determines the first difference for seasonal models in the non-seasonal component.

P determines the order of lags for Autoregressive models in the seasonal component, Qdetermines the order of lags for Moving Average models in the non-seasonal component, and Ddetermines the order of seasonal differencing. 4 after the model is the frequency of the seasonality.

The best model for the data is as it minimizes the ‘aic’ to the maximum among all other models.

The steps that I took to determine the best model are as follows:-

* I used auto.arima to figure out the best model and used options such as trace (to record all the steps), method = ML, and stepwise option is not running in the model.
* Then quite a few combinations of the model ran to figure out the one with the least aic.
* The lowest aic turned out to be of which was 809.05.
* The coefficients and other variables of the model are as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Best Model | ARIMA(1,0,0)(1,1,1)[4] with drift | | | |
|  | ar1 | sar1 | sma2 | drift |
| Coefficients | 0.6929 | 0.4302 | -0.8934 | 17.4752 |
| Standard Errors | 0.0990 | 0.2540 | 0.1530 | 1.9655 |
| AIC | 809.05 | | | |

Table 14

Chart

Description automatically generated

Figure 15

Figure 15 displays the correlogram of the residuals for the best version of . The best version of is which means that non-seasonal part of the model has AR=1, MA=0, and Seasonal = 1. Whereas seasonal part of the model has AR=1,MA=1, and seasonal=1 with a frequency of 4. The residual of the model looks like white noise as all the lags are within the dotted lines (confidence interval) and there is no autocorrelation visible at all.

Chart, scatter chart

Description automatically generated

Figure 16

In the forecast shown in figure 16, there is 95% predictability that the forecast will be within the gray. Looking at the previous data, prediction shown looks pretty accurate but if we compare it with figure 12, then we would see the difference of how much GDP dropped in 2020 so this forecast might be somewhat incorrect.

A picture containing text, antenna

Description automatically generated

Figure 17

Figure 17 is displaying the AirPassengers series into mean and variance stationary graph.