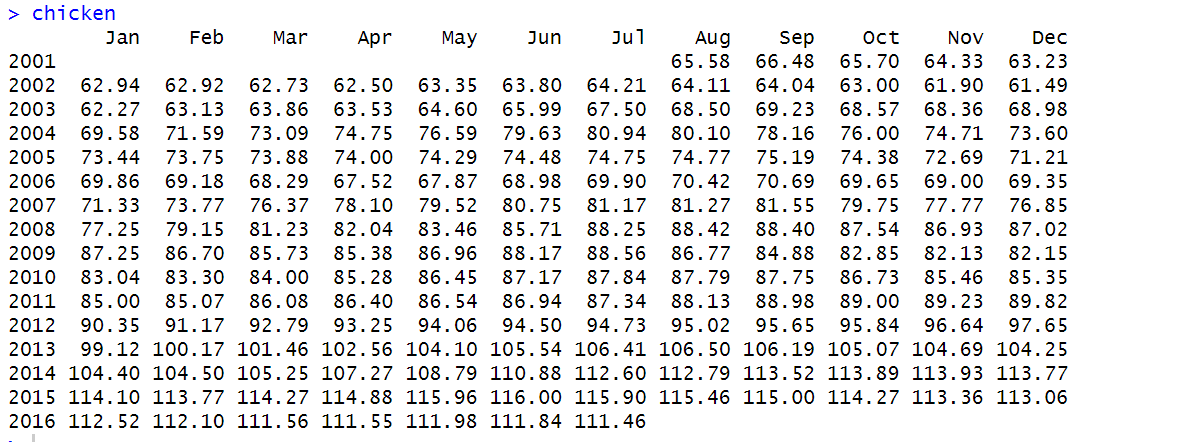
**Manual Assignment #3**

Conduct a thorough analysis of the dataset chicken, available in the astsa package. Your analysis should be typed and should include a number of figures (at least 6, but more may be warranted) that you generate in R yourself. Your analysis should lead to an ARIMA or SARIMA fit (or possibly a discussion of competing models).

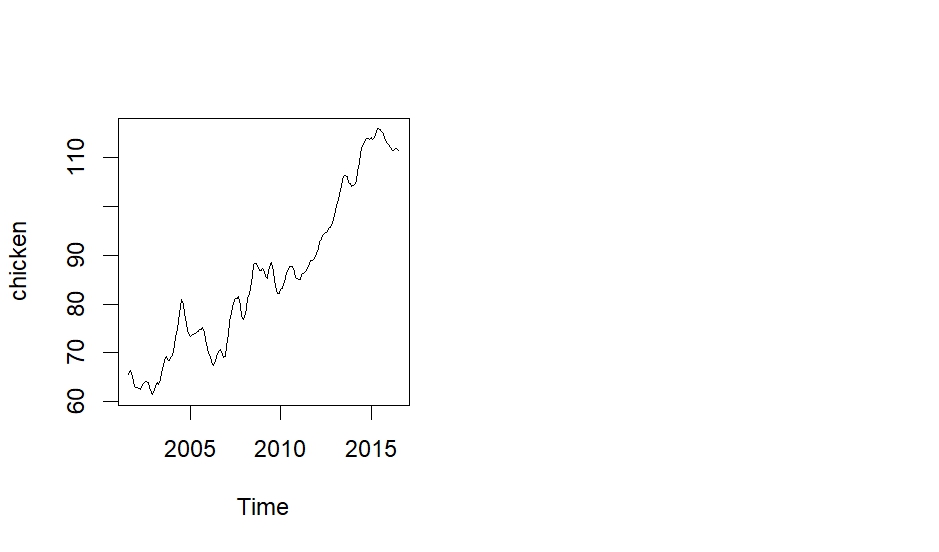
**Solution:**

1. **Information about chicken dataset time series and visual plot:**

Chicken time series contains information about chicken price (US cents per pound) from the month of August 2001 to July 2016. Below is the snapshot of the data in the chicken dataset.



When we plot the chicken dataset, the plot looks like below



1. **Making the chicken time series stationary:**

As we can see from the plot of chicken time series, we can observe a clear upward trend. We do not observe any seasonality or variability in the data, so we difference the data once and then plot the differenced time series.

R code:

par(mfrow=c(1,2))

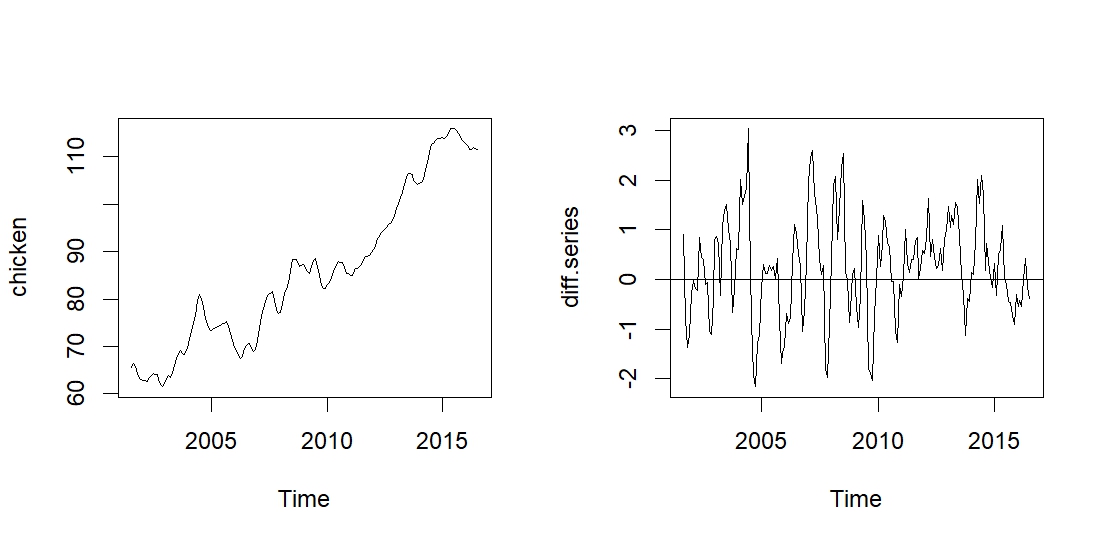
plot(chicken)

diff.series<-diff(chicken)

plot(diff.series)

abline(h=0)

Output:



As we can see from the above 2 plots, left one is the original data and right side one is one differenced data. After apply differencing once, the trend is gone and the series appears to be stationary with mean 0.2563128 (calculated from R -mean(diff.series)) which is very small and so, we can consider the mean to be 0.

1. **Augmented Dickey Fuller Test Analysis:**

Lets run dickey fuller test to check if additional differencing is necessary or not.

Before we run the test, we need to calculate the lag value by using below code:

R code:

(length(diff.series)-1)^(1/3)

Output: 5.625226

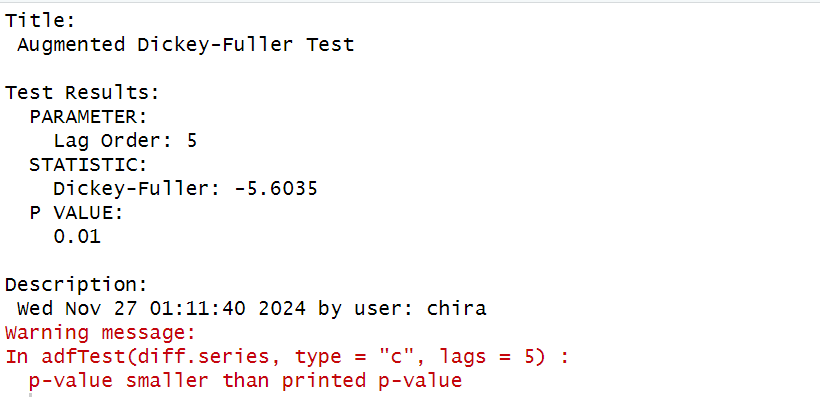
Based on the above value, we consider the value of lag to be 5.

R code:

Case 1: Running the adfTest considering type=’c’ ie constant is present

adfTest(diff.series,type='c',lags = 5 )

Output:

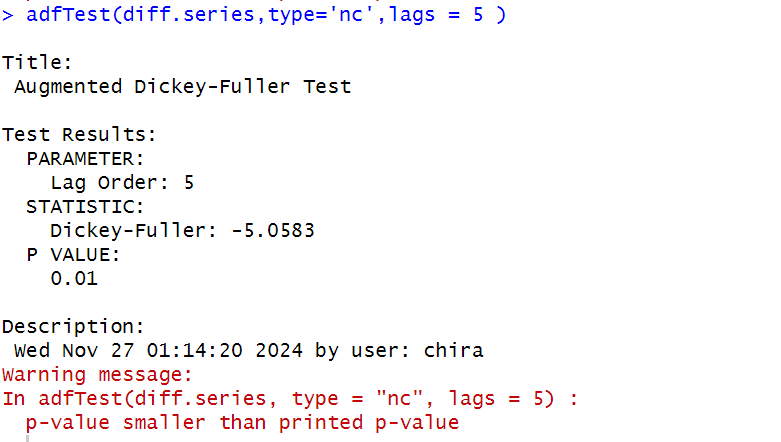


The output of pvalue is 0.01 and the actual p-value is even smaller than the printed value, which means that it is not necessary to difference the data any further.

Case 2: Running the adfTest considering type=’nc’ ie mean is zero

adfTest(diff.series,type='nc',lags = 5 )

Output:



If we consider the differenced time series to be mean zero model, even then the p-value is smaller than the printed value which is 0.01. This means that no additional differencing is necessary.

1. **Sample ACF and Sample PACF plots of the differenced time series.**

We then plot the ACF and PACF plot of the diff.series for initial analysis of the order of auto regressive, seasonal auto regressive, moving average, seasonal moving average and seasonality if present.

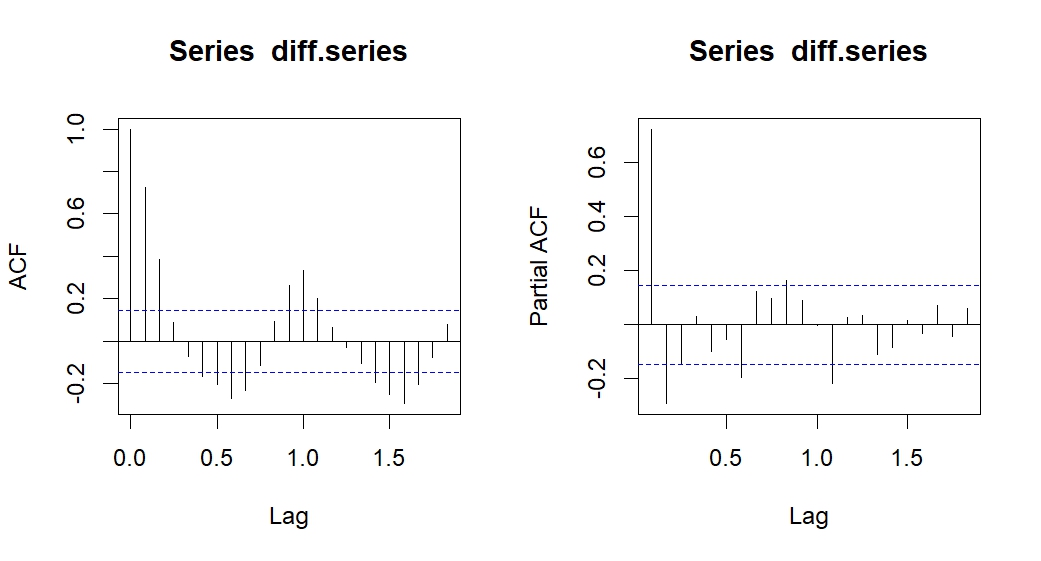
R code:

par(mfrow=c(1,2))

acf(diff.series)

pacf(diff.series)

Output:



ACF and PACF interpretation:

If we consider the ACF plot, there are significant spikes at lag 1 and lag 2, then it slowly decays and goes to the negative part and again rises with major spike at lag 6, lag 7 and lag 8. Again the spikes move in positive direction with significant spike at lag 11 and 12 and similarly at the negative side of lag 18 and lag19. We can see there is an oscillating series of spikes here. There is seasonality present but the spikes are making the interpretation of ACF plot little difficult. It maybe because of the presence of AR component the graph behaves like this. We guess that q and Q are 0 based on the ACF plot.

If we consider the PACF plot, the spikes on the plot are slowly decaying. There is no cut-off. There is significant spike at lag1 and lag2. Then another at lag 7 and lag 12 and then all the spikes are within the threshold line. We can guess the p=2 and P=1 with differencing to be 0 based on the spikes for seasonal and non-seasonal part. The S=12 ie the seasonal value.

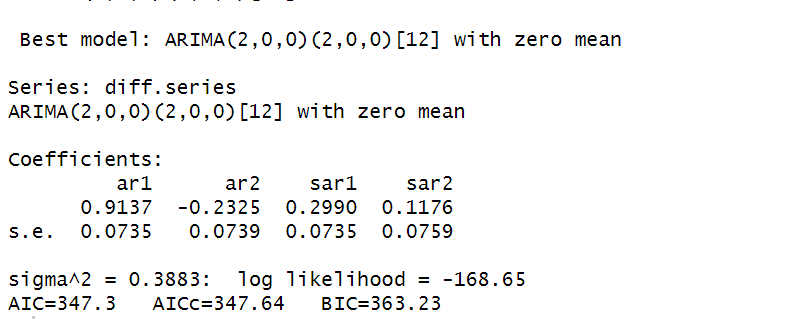
1. **Auto.arima output with trace=TRUE**

We run the auto.arima function in R to check which model it selects for diff.series based on the lowest value of AICc.

R code:

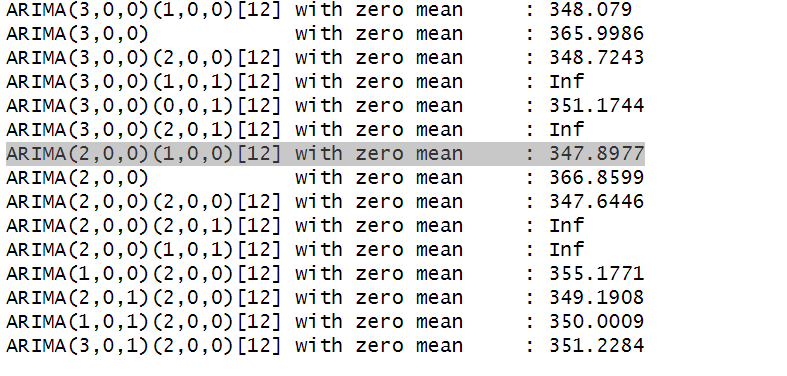
auto.arima(diff.series, ic = "aicc",trace = TRUE, approximation = FALSE)

Output:

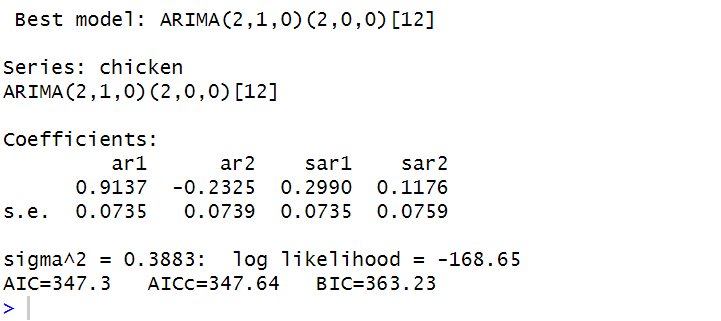


Based on the lowest AICc value the auto.arima model selects ARIMA(2,0,0)(2,0,0)[12] with mean zero.

The model that we had interpreted based on ACF and PACF ie ARIMA(2,0,0)(1,0,0)[12] with mean zero has AICc value as 347.8977, which is not very high from the best selected model.



If we run the auto.arima call on the original dataset, we get below result:



When we run auto.arima on orginal dataset ie chicken, we get the ARIMA(2,1,0)(2,0,0)[12]. The only difference between this model and the differenced model is of 1 lag ie the value of ‘d’ in the non-seasonal part. I think that, since we have done 1 differencing on the chicken dataset, d=1 is not reflecting in the model prediction of differenced dataset. Even the predicted values for AR coefficients and seasonal AR coefficients are same. Let us consider the model selected by auto.arima ie ARIMA(2,0,0)(2,0,0)[12] with mean zero and define the model equation.

1. **Model Equation**

Using the best selected model ARIMA(2,0,0)(2,0,0)[12] with mean zero and the AR and seasonal AR coefficient values estimated by auto.arima(), below is the equation of the fitted model.

Values:

ar1= 0.9137 ar2= -0.2325 sar1= 0.2990 sar2= 0.1176

Equation:

equation

Where equation ie one differenced X\_t.

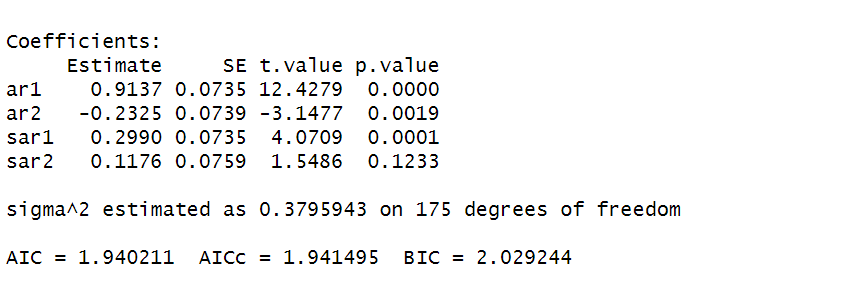
1. **Model Diagnostics by sarima()**

Below is the model diagnostics by sarima() for mean zero model.

R code:

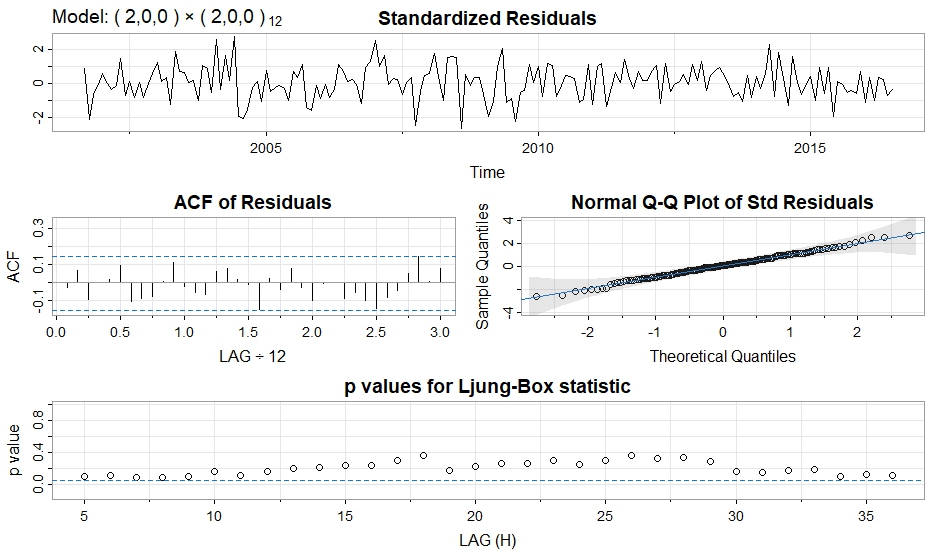
sarima(diff.series,2,0,0,2,0,0,12,no.constant = TRUE)

Output:



The AR and seasonal AR values estimated by sarima function are same as auto.sarima function.

Plot:



Standardized Residual: The residual graph appears to be stationary and revolving around mean 0.

ACF of Residual: All the lags are within the dotted blue line, which makes it consistent with white noise.

Normal Q-Q Plot: All the residuals lie along the line making it normally distributed. There are no departures from normality.

Ljung-Box Statistics: p-values of first few lags lie close to the threshold line. From lag 10 onwards, the p-values are slightly moving away from the threshold line but lie below 0.4 value. This might be troublesome and difficult to conclude whether we can just reject null hypothesis or no for initial few lags.