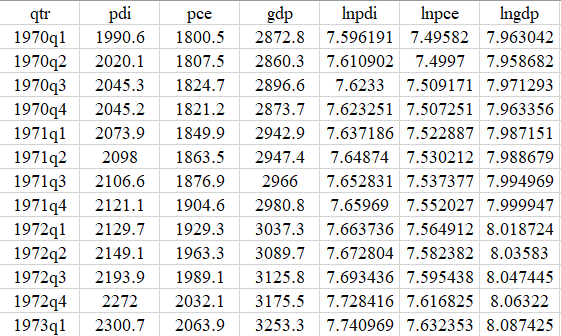
# ASSIGNMENT 2 - ARIMA

**Data used:** Here we will be using a USA GDP data which is quarterly which goes from 1970 to 1991. The following screenshot shows the first few rows of the data. Since the pdi, pce and gdp values are very high, we take the log of these.

****

**Checking for the stationarity**

We are going to consider two parameters namely, lnpce (log of the actual pce values) and lnpdi (log of actual pdi values)

**STEP 1 - IDENTIFICATION**

**How can the appropriate model be identified?**

Since, ARMA/ARIMA is a method among several used in forecasting variables, the tools required for identification are: correlogram, autocorrelation function and partial autocorrelation function. The partial autocorrelation (PAC) measures correlation between (time series) observations that are k time periods apart after controlling for correlations at intermediate lags (i.e., lags less than k). In other words, partial autocorrelation is the correlation between Yt and Yt−k after removing the effect of the intermediate Y’s (measures the marginal impact).

To identify the appropriate ARMA/ARIMA model, I have outlined 5 procedures:

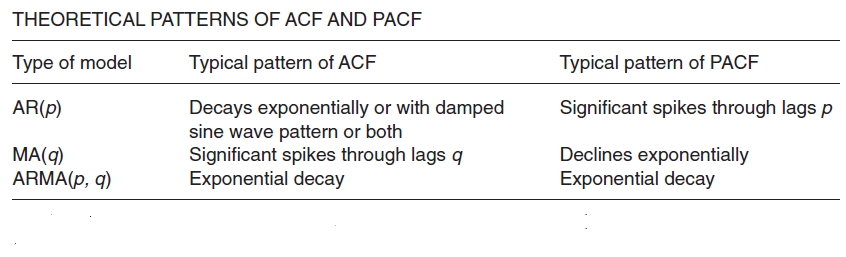
(1) plot the series to visualise if stationary or not.

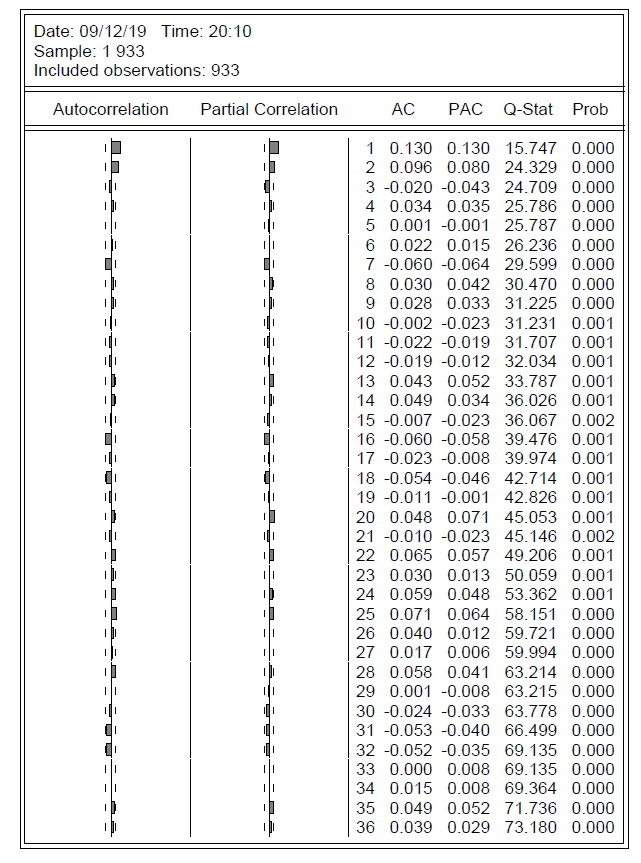
(2) from the correlogram, calculate the ACF and PACF of the raw data. Check whether the series is stationary or not. If the series is stationary go, to step 4, if not go to step 3.

(3) take the first differences of the raw data and calculate the ACF and PACF from the correlogram.

(4) visualise the graphs of the ACF and PACF and determine which models would be good starting points.

(5) estimate those models. Using EViews10.

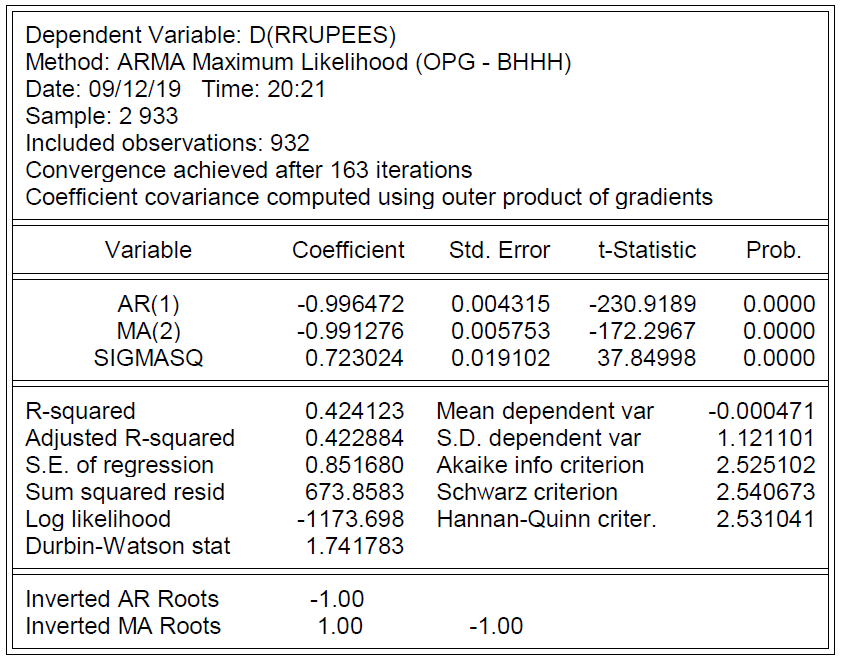
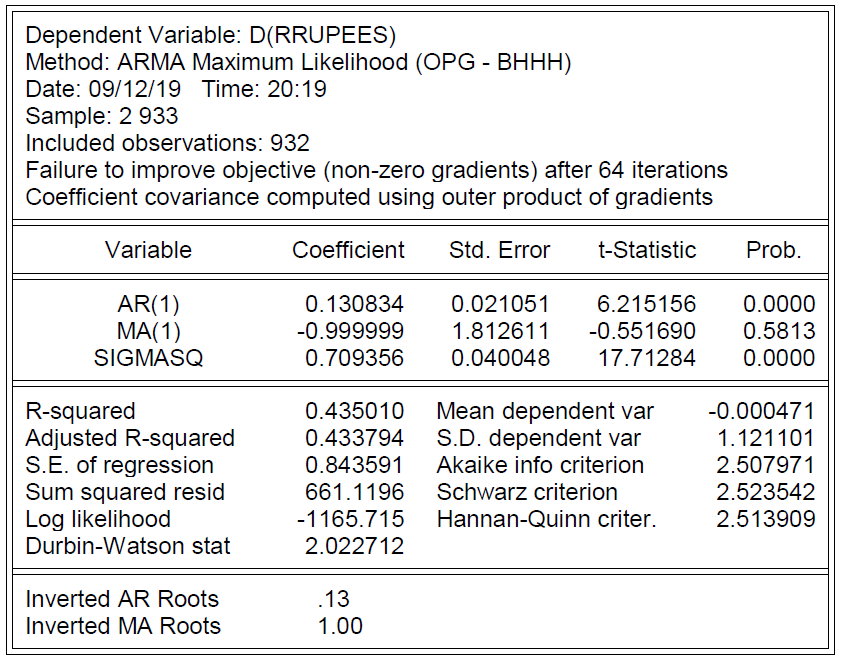




From the above plot we can select the values of “p” and “q” by looking at the spikes, and select the models. For example, from the above correlogram, the ACF and PACF values at lag 1 and 2 are significant. Also, both the ACF an PACF plots follow the same pattern, meaning we have an ARMA model. So from the above knowledge we can have models ARMA(1,0,1), ARMA(1,0,2), ARMA(2,0,1) and ARMA (2,0,2). The next step is where we find which model is most significant out of all the listed models.

**STEP 2 – ESTIMATION**

**Different plots for different models such as shown below.**

****

Looking at the above plots, our most significant model out of all the models would be one satisfying the following criterion:

1. **Model that has the most significant coefficients.**
2. **Model that has the least volatility.**
3. **Model that has lowest value for AIC and BIC.**
4. **Model that has highest values for R-square and Adjusted R-square.**

Consider the following sample table, as shown in the table we have listed out the values for our variables which would help us to determine the most appropriate model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **ARIMA(1,1,1)** | **ARIMA(1,1,8)** | **ARIMA(8,1,1)** | **ARIMA(8,1,8)** |
| **Significant coeff** | 0 | 2 | 2 | 0 |
| **Sigma ^2** | 1175.38 | 1039.09 | 1078.35 | 1137.33 |
| **Adj R-square** | 0.059 | 0.169 | 0.137 | 0.09 |
| **AIC** | 10.00 | 9.89 | 9.93 | 9.99 |
| **SBIC** | 10.12 | 10.01 | 10.04 | 10.10 |

From the above table, ARIMA(1,1,8) looks like the most appropriate model, as it has most number of significant coefficients, lowest volatility, highest R-square and Adjusted R-square, and lowest SIC and BIC.

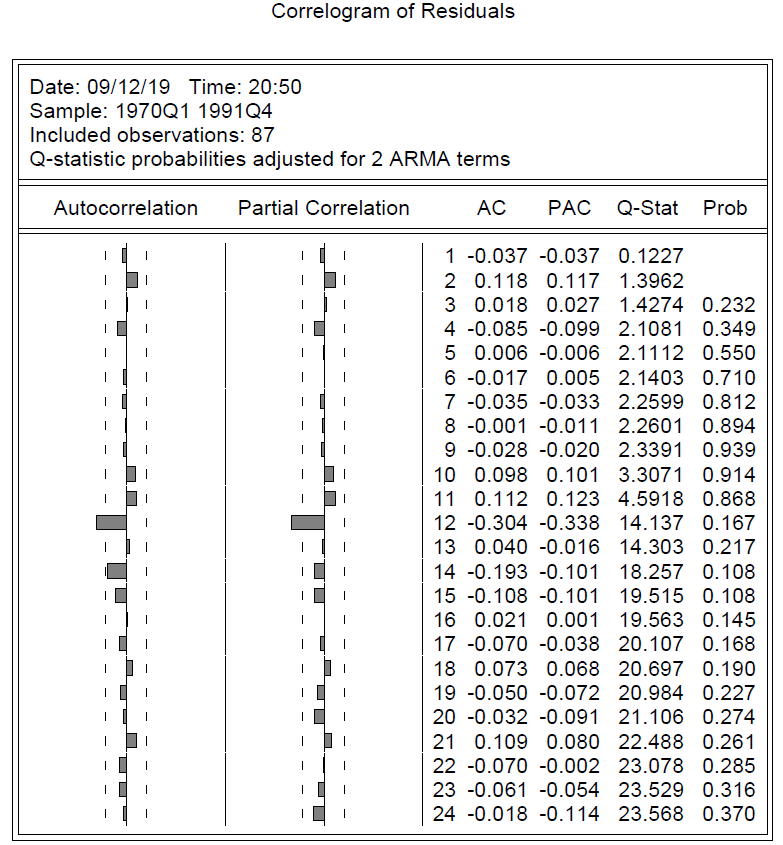
**STEP 3 – DIAGNOSTIC CHECKING**

How do we know that the model chosen above is a reasonable fit to the data? One simple diagnostic is to obtain residuals, and obtain the ACF and PACF of these residuals, say, up to lag 25. The estimated AC and PACF are shown in Figure 22.5. As this figure shows, none of the autocorrelations

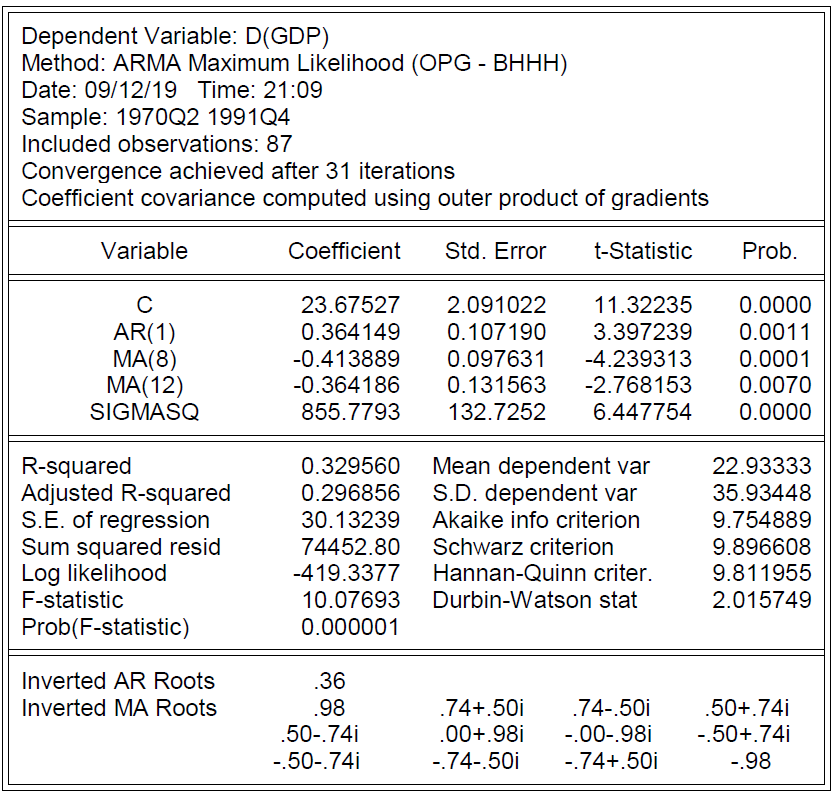
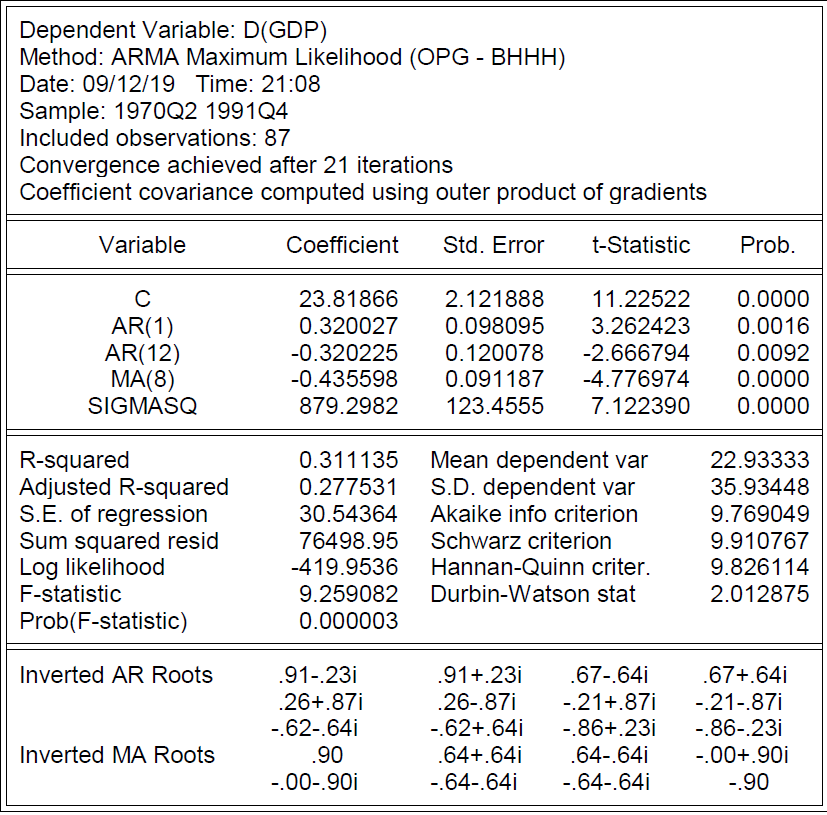
Having identified ARIMA(1,1,8) as the most appropriate model the next step is to perform some diagnostics to make sure that there is still no information in ARIMA(1,1,8) that is left uncaptured. Now in order to know we would need to plot the **correlogram of residuals**. The below figure shows the correlogram for the residuals for the ARIMA(1,1,8) model.

To summarize, following are the steps to take if the we still find some significant lag in the residuals

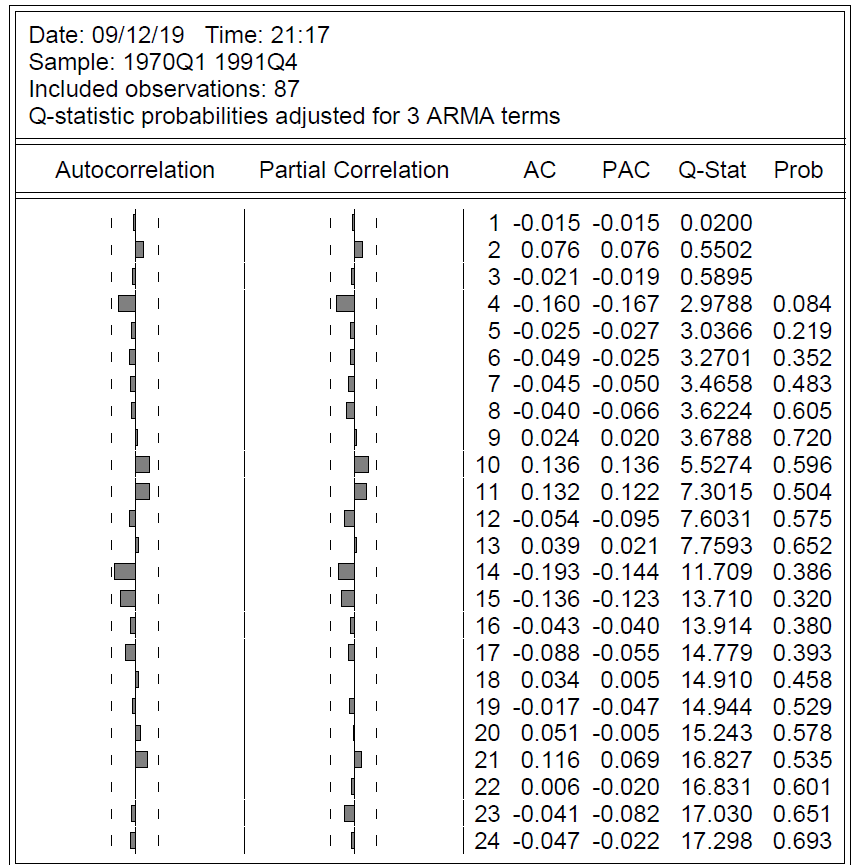
1. Find the most appropriate model
2. Check the residuals correlogram to see if there’s any information yet to be captured in the model.
3. A flat correlogram is the most ideal.
4. If a lag is significant, re-estimate the model.
5. Check the residuals correlogram again.
6. But, avoid over-fitting a model.
7. Perform Ljung-Box text for squared residuals (autocorrelation test).



To cover the remaining information from the model, we will include lag 12 as well in our ARIMA(1,1,8) model. Once that is done, we will again check the residuals and look at the correlogram. The following statistics shows the stats for the models after inclusion of lag 12.

****

As we can see from the below correlogram for the residuals, our spikes are flat and they lie within the 95% confidence interval. Now we can safely say that we have captured all the relevant information from our model. We can see that all the lags for both ACF and PACF lie within this confidence interval. So our forecast will be based on this model.

****