```
### Lecture AR-MA: R commands used
> require(quantmod)
> getSymbols("UNRATE",src="FRED")
[1] "UNRATE"
> dim(UNRATE)
[1] 831   1
> head(UNRATE)
           UNRATE
1948-01-01    3.4
1948-02-01    3.8
1948-03-01    4.0
1948-04-01    3.9
1948-05-01    3.5
1948-06-01    3.6
> rate <- as.numeric(UNRATE[,1])
> ts.plot(rate)
> m1 <- ar(rate,order.max=15)  ## AR order selection using AIC
> m1$order
[1] 13
> m2 <- arima(rate,order=c(13,0,0))
> m2
Call:
arima(x = rate, order = c(13, 0, 0))

Coefficients:
         ar1     ar2      ar3      ar4     ar5      ar6      ar7
ar8
      1.0045  0.1911  -0.0594  -0.0457  0.0399  -0.1287  -0.0432
0.0431
s.e.  0.0346  0.0488   0.0492   0.0491  0.0493   0.0493   0.0495
0.0493
          ar9     ar10    ar11     ar12    ar13  intercept
      -0.0040  -0.0824  0.1145  -0.1642  0.1211     5.6337
s.e.   0.0493   0.0492  0.0494   0.0490  0.0348     0.4787

sigma^2 estimated as 0.0366:  log likelihood = 192.75,  aic =
-355.49
> tsdiag(m2,gof=36)
### Model refinement
> c1 <- c(NA,NA,0,0,0,NA,0,0,0,NA,NA,NA,NA,NA)
> m3 <- arima(rate,order=c(13,0,0),fixed=c1)
Warning message:
In arima(rate, order = c(13, 0, 0), fixed = c1) :
  some AR parameters were fixed: setting transform.pars = FALSE
> m3
Call:
arima(x = rate, order = c(13, 0, 0), fixed = c1)

Coefficients:
         ar1     ar2  ar3  ar4  ar5      ar6  ar7  ar8  ar9     ar10
ar11
      0.9992  0.1406    0    0    0  -0.1528    0    0    0  -0.0718
0.1169
s.e.  0.0340  0.0404    0    0    0   0.0264    0    0    0   0.0410
```

```
0.0487
         ar12    ar13  intercept
      -0.1744  0.1287     5.6368
s.e.   0.0488  0.0345     0.4738

sigma^2 estimated as 0.0368:  log likelihood = 190.42,  aic =
-362.84
####
The AIC reduces to -362.84 ndicating that it is ok to set those
parameters to
zero (as given in c1).

### Use ARIMA models
> require(forecast)
> auto.arima(rate)
Series: rate
ARIMA(2,1,2)

Coefficients:
         ar1      ar2      ma1     ma2
      1.6537  -0.7759  -1.6319  0.8487
s.e.  0.0414   0.0459   0.0408  0.0474

sigma^2 estimated as 0.03831:  log likelihood=177.6
AIC=-345.19   AICc=-345.12   BIC=-321.59
> m4 <- arima(rate,order=c(2,1,2))
> m4
Call:
arima(x = rate, order = c(2, 1, 2))

Coefficients:
         ar1      ar2      ma1     ma2
      1.6537  -0.7759  -1.6319  0.8487
s.e.  0.0414   0.0459   0.0408  0.0474

sigma^2 estimated as 0.03813:  log likelihood = 177.6,  aic =
-345.19
> tsdiag(m4,gof=36)
### The model checking result shows some serial correlations remain
in
### the residuals.

############# Model comparisons
Two types of comparison.
A). In-sample fit.
  Here we select one of the criteria, say, AIC.
  We select the model with smaller AIC as the preferred model.
For unemployment rate, model "m3" has AIC = -362.84.
On the other hand, model "m4" has AIC = -345.19.
So model "m3" is selected.

B). Out-of-sample comparison.
 Here we divide the data into two subsamples: training sample and
forecasting sample.
```

This is done by selecting a forecast origin. For illustration below, I used t = 770 as the forecast origin.

We then apply "backtest", which uses a rolling of estimation-prediction to compute 1-step ahead forecasts for a given model, starting with forecast chosen forecast origin. Specifically, we estimate the model using the first 770 data points, then forecast 771 to compute forecast error. We then re-estimate the model using the first 771 data points and predict t=772 to compute forecast error. This procedure is repeated until the end of the sample. In unempolyment example, we estimate the model using 830 data points, then predict t=831. The forecast errors can be used to compute mean squared forecast error (MSE) and mean absolute forecast error (MAE).

This procedure is applied to each of the competing models. If MSE is used, one selects the model with the smallest MSE as the preferred model.

The "backtest" can be done using the R script "backtest.R", available from the course web. You need to save the script into your working directory of R. Then, you can use the command to perform the rolling estimation-prediction method. See the demonstration below:

```
> source("backtest.R")
> backtest(m3,rate,770,fixed=c1)
[1] "RMSE of out-of-sample forecasts"
[1] 0.1455117
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.118224
There were 50 or more warnings (use warnings() to see the first 50)
> backtest(m4,rate,770)
[1] "RMSE of out-of-sample forecasts"
[1] 0.1482988
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.1180253
>
### From the output, model "m3" is preferred based on MSE.
But "m4" is preferred based on MAE.
```

Discussion: A drawback of "backtest" comparison is that the result may depend on the forecast origin. One should use sufficient number of data points in the forecasting subsample to obtain more reliable comparison.