HW 4 (Due April. 22, 2015)

1. Suppose $\{X_t\}$ is an MA(2) process, $X_t = W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2}$, $\{W_t\} \sim \mathrm{WN}(0, \sigma^2)$. If the AR(1) process

$$(1 - \phi B)X_t = Y_t$$

is mistakenly fitted to $\{X_t\}$, determine the autocovariance function $\{Y_t\}$.

2. Download the txt file of CO_2 series from Mauna Loa, Hawaii.

   (a) Use the codes

```
X=scan("co2.txt")
X=matrix(X,,2,byrow=TRUE)
T=X[,1]
X=X[,2]
plot.ts(X,col="blue")
```

   to plot the series

   (b) You can see the plot has a trend which is increasing in time t and also have a seasonal component. Thus, this time series $\{X_t\}$ can be modeled as

$$X_t = m_t + s_t + Y_t.$$

   I suggest using a quadratic function $a + bt + ct^2$ to fit the trend m_t. Using the OLS method to obtain $\widehat{a}, \widehat{b}, \widehat{c}$; i.e., OLS estimators of a, b, c, respectively. What do you think about the fit?

   (c) Now take the residual $\widehat{X}_t = X_t - \widehat{m}_t$, where $\widehat{m}_t = \widehat{a} + \widehat{b}t + \widehat{c}t^2$. Plot the sequence, its ACF and PACF to lag= 100. What do you think about the residuals?

   (d) After m_t, we now think about s_t. This is a yearly data; in each year we have 12 observations. So we can set the period $s = 12$; i.e., $s_t = s_{t+12}$ for any t. Thus, the unknown seasonal components will be $\{s_1, \ldots, s_{12}\}$. Obtain point estimator $\{\widehat{s}_j; j = 1, \ldots, J\}$. (Just use sample average for each j). Now plot the series of $\widehat{Y}_t = X_t - \widehat{m}_t - \widehat{s}_t$ and its ACF and PACF to lag= 100. What do you think about this sequence?

   (e) Take one order of differencing; i.e., $\nabla \widehat{Y}_t$. Plot the series, its ACF and PACF to lag= 100. What do you think about this sequence? Run the Dickey-Fuller test to see whether this differencing is needed or not.

   (f) Now instead of using deterministic s_t, using SARIMA model (R codes: *sarima*) to fit the data (or to capture the seasonal component stochastically). Find the optimal SARIMA model you prefer for the dataset. Explain why you preferred it.