STAT 443: Time Series and Forecasting

Lab 4: SARIMA Processes

- The lab must be completed in `R Markdown`. Create a **pdf file** (via 'knit to pdf') and use it as your lab submission.

We have seen both autoregressive (AR) and moving average (MA) models, as well as hybrids of these known as autoregressive moving average (ARMA) models. While these are quite versatile models, recall that they are used to model series that are *stationary*.

1. Without using any mathematical notation, describe in words what it means for a time series to be stationary.

2. Consider a realization of a process with values given in file `lab4data.csv`. Read in the data, coerce it into a `ts` object, plot and comment on whether the series appears to satisfy the requirements of stationarity.

3. One common way of removing a trend is to difference the data, where instead of looking at the time series $\{x_t\}$ we look at $\{y_t\}$ with

   $$y_t = \nabla x_t = x_t - x_{t-1}.$$

   (Note this series will have one less term than the original.) Using the data above, determine and plot the differenced time series $\{y_t\}$. Comment on the resulting plot and the acf of $\{y_t\}$.

   A useful function here is `diff(x, lag=1, difference=1)`, which returns suitably lagged and iterated differences. Use R help to learn about options `lag` and `differences`.

4. In order to remove a seasonal effect, we could difference over the seasonal period. For instance, take our de-trended data $\{y_t\}$ and difference again but at a lag equal to the seasonal period, s. The new series will be

   $$\nabla_s y_t = y_t - y_{t-s},$$

   where s is the period of the seasonal effect. Note the series will again become shorter, this time with s fewer terms. Choosing the appropriate value of s here, apply seasonal differencing to the series from part 3 and plot the resulting series. Plot the acf of $\nabla_s y_t$. Does your $\nabla_s y_t$ resemble white noise?

5. Suggest which type of model from the SARIMA family you would use for the original data.

6. (**Theoretical exercise**) We have seen that removing trends and seasonality can be as simple as differencing successively at different lags. In such cases, suitable models are *integrated* ARMA (ARIMA) models, or SARIMA if we include seasonal differencing. One difficult aspect of these models is describing them in mathematical terms. Let us try re-using the notation we used with ARMA models to describe these processes.

(a) Firstly, we took our original time series X_t and differenced it to obtain $Y_t = X_t - X_{t-1}$, then differenced again at a longer lag to obtain $W_t = Y_t - Y_{t-s}$. Combine these two operations to express our final model W_t in terms of the original series X_t.

(b) Recall the differencing operator B which takes a series and returns the series at lag 1. For instance, $BX_t = X_{t-1}$. Express the differenced series Y_t in terms of B and X_t.

(c) Now if $B^s Y_t = Y_{t-s}$, express your answer from (a) in terms of B. That is, write W_t in terms of B, s, and X_t.