Topics of the lab: Fitting ARIMAX models.

In this lab we are going to look at the procedure of fitting regression models with ARIMA errors, that is, we'll consider models of the form

$$Z_t = \beta_0 + \sum_{i=1}^{m} \beta_i X_{i,t} + \varepsilon_t,$$

where the process ε_t is an ARIMA process. More precisely, we are going to assume that the error process ε_t is stationary (meaning that usually we have to fit the model to suitably differenced versions of Z_t and X_t, if they are not stationary).
The model fitting process is as follows:

i) Fit a linear regression model for Z_t by using $X_{it},\ i = 1, \ldots, m$ as independent variables.

ii) Determine the form of a possible ARIMA (or SARIMA) model for the residuals

iii) Fit corresponding ARIMAX model to the original data and test the independence of residuals. Go back to step ii) if necessary.

Exercises:

0) Load the time series data from the file `lab12.txt` (available from Moodle).

1) Suppose that we know that good regressors for $Z_{1,t}$ should be $X_{1,t-6}$. In order to fit an ARIMAX model, the series should have the same length (and aligned correctly). Since we do not have regressors for the first 6 observations of the series z1, we should leave those values out. This can be done with the command

   z11=**window**(Z1, **start**=7).

   Similarly, the last 6 values of x1 should be removed from the regressor series. Please find a suitable ARIMAX model for z1, compare this model to the best ARIMA model for z1 and also find predictions for the next 6 observations by both models.

2) Consider finding an ARIMAX model for z2 by using x21 of lag 4 and a x22 of lag 2 for regressors. Note that two regressor series should be specified as xreg=**data.frame**(x1,x2) inside the arima() command. When predicting future values of z2, one should specify the new values of regressor series by newxreg=**data.frame**(x1=x1**_new**,x2=x2**_new**). Find a suitable ARIMAX model and predict the next two values of z2.

3) Find cross correlations between z1 and x1. Do you understand the meaning of cross correlations? Do you see, why it was reasonable to use x1 of lag 6 as the regressor for z1? Find also cross correlations between z2 and x21, between z2 and x22.

There are no reasons why we should use only one lag of the regressor series for predicting a future value of a series of interest. We can easily use different lags of a given regressors series as different regressor series of an ARIMAX model. The main question is how to select meaningful lags.
We saw that cross correlations may give some indication about which lag may be the most meaningful one but it may be quite difficult to decide how many different lags to use. It turns out that so called prewhitening technique may help us to make such decisions.
The steps are as follows:

1. Fit a model to the series X_t, find it's residuals A_t
2. Use the same model (with exactly the same parameters) for predicting values of Z_t, compute the residuals B_t
3. Find the cross correlations between A_t and B_t. If there are only a small number of cross correlations and those are on positive side, then use corresponding lags of X_t in the ARIMAX model for Z_t

Exercises

4) Use the prewhitening idea to select best lags of the series x5 for predicting values of z5. Fit corresponding ARIMAX model and predict as many future values as there are available values for the needed lags of x5. Hint: if the model fitted to x5 is m5x, then the same model can be used for predicting values of z5 by using the command

```
m5z=arima(z5,order=same_as_for_m5x,fixed=m5x$coef)
```