

Week 12: Arima and Sarima

Devesh Tiwari

July 27, 2016

Introduction

Up until now, we have discussed how to model weakly stationary time series data. This week, the Async. materials covered how to model data that are not stationary in the mean (but stationary in the variance). This procedure requires us to difference the data (using the diff command) until the data are stationary in the mean. After that, you can use ARMA to model the data. Relatedly, we can now handle seasonal effects (sarima) models.

Notes on the modeling process

When it comes to building a model and reporting your results, whether it is a CLM or a time series model, it is important to be systematic and to be able to justify your modelling decisions. These decisions take place at different stages of the modeling process (EDA, post regression diagnostics), but it is important that you can justify (in words) why you made the decisions that you made.

For time series data:

- (1) Remember that the ARMA models should only be used on weakly stationary data. So your first task is to assess whether or not your data are stationary in the mean and variance. Often times, this is done visually through EDA.
- (2) Assessing stationarity also means being able to assess whether or not there are seasonal effects and whether they should be differenced out.
- (3) Keep track of your differencing and if you differenced the nonseasonal component (d) or the seasonal component (D).
- (4) Once you have a stationary series, find the right model. To do this, start with some more EDA (acf and pacf) for ideas, but try several models. You have d, and D from steps 2 and 3, now you have to find p,q,P,Q.
- (5) Once you have some candidate models (using whatever criteria you like), examine the residuals to make sure they resemble white noise.
- (6) If you are interested in forecasting, compare the candidate models using an out of sample procedure.

The most important things to keep in mind are:

- (a) Avoid output dumps (make sure every chart has a purpose)
- (b) Your modeling choices are justified

Sarima example

- (1) Examine the dataset *cmort*.
- (2) Do you think this is a weakly stationary time series?
- (3) Do you think that there are seasonal effects?
- (4) Using your EDA as a guide, what is your best guess as to what model is appropriate?
- (5) Fit that model and other models of interest and compare.

Helpful commands:

`diff(cmort, lag = N)` # Creates the seasonal difference
`diff(cmort, diff = N)` # Creates the N-th difference

`acf(X, lag.max = Z)` # acf of X where you can set how far in the past you want to look.

`modelName <- Arima(cmort, order = c(p,d,q), seasonal = list(order = c(P,D,Q), period = M), method = "ML")`