INDIAN INSTITUTE OF TECHNOLOGY MADRAS
Department of Chemical Engineering
CH 5350 Applied Time-Series Analysis

Project Statement

Due: Monday, December 01, 2014 11:00 PM

---

1. [Threshold Auto-regressive Model]

A threshold auto-regressive (TAR) model is an extension of the classical AR(P) model to describe non-linear and non-stationary processes. It is, essentially, a multiple-linear model with r regimes determined by thresholds, $\gamma_i,\ i = 1, \cdots, r-1$ and a delay $d \leq P$. A two-regime TAR(2,$d = 1$) model, for instance, has the form:

$$v[k] = -d_0^{(1)} - d_1^{(1)} v[k-1] - d_2^{(1)} v[k-2] + e_1[k], \qquad v[k-1] \leq \gamma,\ e_1[k] \sim \text{i.i.d.}(0, \sigma_1^2) \tag{1a}$$
$$v[k] = -d_0^{(2)} - d_1^{(2)} v[k-1] - d_2^{(2)} v[k-2] + e_2[k], \qquad v[k-1] > \gamma,\ e_2[k] \sim \text{i.i.d.}(0, \sigma_2^2) \tag{1b}$$

The **TSA** package offers routines to simulate, estimate and test TAR models, through `tar.sim`, `tar` (different from `tar` in the **stats** package) and `tsdiag.TAR`, respectively. Predictions are performed by `predict.TAR`. See the **TSA** package for more details.

(a) For the data set given in `projq1a.Rdata`, build a suitable two-regime TAR model. Estimate the coefficients, threshold and delay. Note that the routine uses the notation $\phi_i^{(j)} = -d_i^{(j)}$ and p_1, p_2 for the orders in each regime, respectively. The final model that you report should have been subjected to all critical diagnostic tests (residual tests, parsimony, etc.).

(b) The standard errors returned by the `tar` routine are based on asymptotic expressions. Perform **bootstrap** simulations to determine the standard errors in each of these estimates. The main difference between the Monte-Carlo and *bootstrap* simulations is that the former uses the true distribution (or the model), whereas the latter solely uses the available data for generating realizations. Use the following procedure for bootstrapping[1]

i. Generate residuals from the obtained "best" model. Denote these by $\varepsilon[k],\ , k = 1, \cdots, N$

ii. Re-sample the residuals with replacement to obtain a new realization of the same length. Call this series $\{\varepsilon^{(i)}[k]\}$. Use the `sample` routine for this purpose. A sample usage is: `epskr1 <- sample(epsk,size=N,replace=T)`.Add this to the predictions of your TAR model to generate a new *artificial* realization of the series, i.e., $\hat{v}^{(i)}[k] = v[k] + \varepsilon^{(i)}[k]$.

iii. Re-estimate your model coefficients (of the same orders and delay) using $\{v^{(i)}[k]\}$

iv. Repeat steps (ii) - (iii) for a desired number of realizations R (e.g., $R = 200$). Compute mean, variance and distribution of the respective coefficients from their estimates across realizations, and compare your answers with those from the `tar` routine.

Note: Students are most welcome to use other bootstrapping procedures for time-series models that they are familiar with.

2. Build a suitable (SARIMA) model for the `retail` data in the TSA package. Report the results from all the important steps and justify your final choice of model.

---

[1]This is a standard, but somewhat elementary, procedure. There exist other sophisticated bootstrapping methods.