SMM634 Coursework 3 2018/2019

Deadline 16 April 2019

NAME:

1. The dataset `timesim.dat`, available on Moodle, contains the following four simulated time series.

$$
\begin{aligned}
Y_t &= E_t - 0.5E_{t-1}, & E_t \sim N(0,1) i.i.d., \quad E_0 = 0 \\
Y_t &= Y_{t-1} + E_t, & E_t \sim N(0;1) i.i.d., \quad Y_0 = 0 \\
Y_t &= 0.5Y_{t-1} + E_t, & E_t \sim N(0;1) i.i.d., \quad Y_0 = 0 \\
Y_t &= Y_{t-1}E_t, & E_t \sim U(0.95;1.05) i.i.d., \quad Y_0 = 1
\end{aligned}
$$

   Use a time series plot to decide whether or not these processes are stationary. If a process is not stationary, suggest a simple method to make it stationary, and try this method out in R. [Marks: 10]

   R-Hint: `timesim[,1]` is the first time series.

2. Consider the AR(3) model with coefficients $\alpha_1 = 0.5$, $\alpha_2 = -0.4$ and $\alpha_3 = 0.6$:

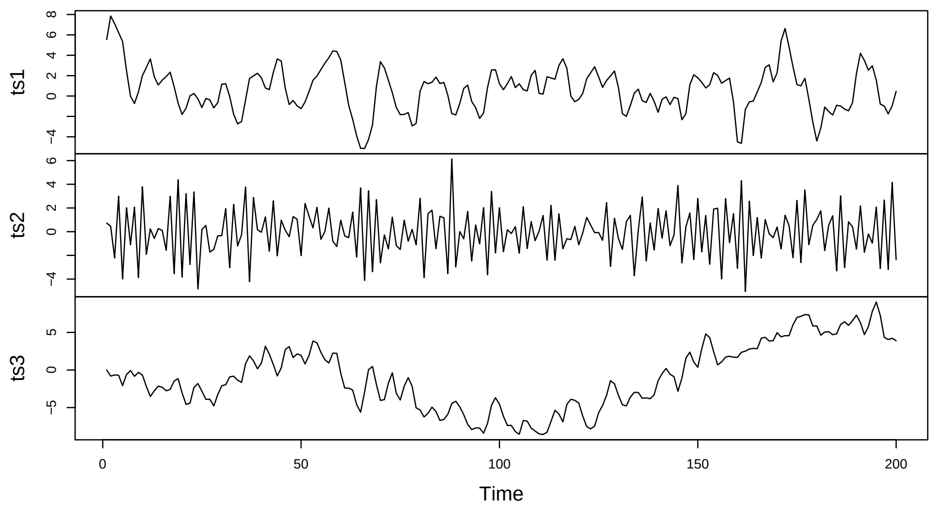
$$X_t = 0.50.5X_{t-1} - 0.4X_{t-2} + 0.6X_{t-3}$$

   (a) Simulate 1 realization of length 100 of this time series and plot it. Does the time series look stationary? [Marks: 3]

   (b) Have a look at (partial) autocorrelations. Do they look as you expect? Comment. Also compare the estimated autocorrelations to the true ones. [Marks: 4]

   (c) Simulate realizations of this time series of different lengths (shorter and longer than 100) and compare each of their estimated autocorrelations with the real ones. What do you observe? [Marks: 3]

3. For each of the following three time series stored in `ARIMAsim2.dat`, find a suitable ARIMA(p,d,q) model and estimate its parameters. [Marks: 10]



4. (a) Generate a simulated data set with $n = 1000$ from the following statistical model

$$Y = X - 2X^2 + X^3 + \epsilon, \quad \epsilon \sim N(0,1), \quad X \sim N(2,2).$$

[Marks: 2]

(b) Using the `cv.glm()` function from the `boot` package compute the LOOCV errors that result from fitting the following two models using least squares:

i. $Y = \beta_0 + \beta_1 X + \epsilon$

ii. $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon.$

[Marks: 2]

(c) Write an R function that calculates

$$CV_{(n)} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{1 - h_{ii}}\right)^2,$$

where y_i and $\hat{y}_i$ are the i^{th} actual and fitted values, respectively, from the original least squares fit, and h_{ii} is the i^{th} diagonal element of the hat matrix. [Marks: 3]

(d) Using your function compute the LOOCV error that results from the models fitted in (b) and compare with the results obtained using `cv.glm()`. Write your findings. [Marks: 3]