The deadline for this project is **May 27th 2023**.

You will upload in the MS Teams Assignment **only one pdf** file (the written document + the code added at the end). Do not send it by email! The pdf will be verified with Turnitin.

Please include your name in the file name.

Requirements:

1. Choose a time series (**any real data**, except for the The quarterly U.S. Gross National Product gnp and AirPassengers presented in Lab4). Look careful at it and make sure that it does not look like any of the two from the lab.

   – Build a model for the time series that you choose – explain all the steps that you go through.

   Important: you cannot use any ready-made function that receives the time series as input and provides the best model as output (like a black box). Instead, you can use the functions from the Labs, or similar ones.

   – Use that model to forecast the next 20 values in the time series and plot them together with the initial time series.

2. In C5 9th slide, you find a suggestion to choose a preliminary value of *p* for an AR(*p*) model. It requires the implementation of the Durbin-Levinson (D-L) algorithm, where you will replace population moments γ(h) by the sample moments $\hat{\gamma}(h)$ (2nd slide C5).

   Take an AR(*p*) model (let's say AR(2) with the parameters Φ =(1,-.9)) and generate 200 observations *x*.

   Then use *x* to estimate the order of the model, as suggested in C5. Compare then the values of Φ_{mm} from the D-L algorithm with the PACF values returned by the function implemented in the software package that you use (e.g., the pacf function in R).

The project will consist of two parts:

1. implementation (preferably in R, but any other language is also allowed). **Attention:** you should not insert any comments that explain your code; only for reading/writing/plotting data, you are allowed to re-use code from internet (e.g., from Stack Overflow) – but in this case, you must clearly indicate the lines that you took and insert a comment with the link of the source.
2. a written document, where you mention the steps that you follow: describe the datasets, plot the time series, interpret the results, plot different useful graphs, explain the choices that you make when building the model, argue why the model is the best (according to which criteria); the goodness of fit analysis; comparison

between the output of your D-L implementation in S2 and the function implemented in the software package etc. (between 2-10 pages).

Attention: you must indicate the source of the time series that you model. Also, all the resources (on internet, books) that you consult for the project must be cited.

You must work individually for the project.

You must present your project in order to be evaluated. The presentations will be scheduled in the last week of the semester. The planning will be available later on Moodle and MS Teams (max 60 minutes for each student).

If you have questions, my email is cidota@fmi.unibuc.ro

Marina Cidota

21.04.2023