PSTAT 174/274 Spring 2023: Homework 5

Note: $\{Z_t\} \sim WN(0, \sigma_Z^2)$ denotes white noise.

1. Create a glossary of R-commands for time series. It should contain all commands that you learned so far in the labs, doing homework, and reviewing posted lecture slides. At the minimum, the glossary should include commands that allow
– define working directory;
– read and plot data;
– simulate and plot ARMA models;
– add trend and mean line to the original time series plot;
– calculate and plot theoretical acf/pacf for ARMA models;
– calculate and plot sample acf/pacf;
– check whether a particular model is causal/invertible (R commands to find and plot roots of polynomials)
– perform Box-Cox transforms;
– perform differencing data at lags 1 and 12;
– perform Yule-Walker estimation and find standard deviations of the estimates.
– perform MLE and check AICC associated with the model.

2. Choose a dataset that you will be interested to analyze for your class final project. URLs of time series libraries are posted on Gaucho Space. Provide the following information about the project:

(a) Data set description: briefly describe the data set you plan to use in your project.
(b) Motivation and objectives: briefly explain why this data set is interesting or important. Provide a clear description of the problem you plan to address using this dataset (for example to forecast).
(c) Plot and examine the main features of the graph, checking in particular whether there is (i) a trend; (ii) a seasonal component, (iii) any apparent sharp changes in behavior. Explain in detail.
(d) Use any necessary transformations to get stationary series. Give a detailed explanation to justify your choice of a particular procedure. If you have used transformation, justify why. If you have used differencing, what lag did you use? Why? Is your series stationary now?
(e) Plot and analyze the ACF and PACF to preliminary identify your model(s): Plot ACF/PACF. What model(s) do they suggest? Explain your choice of p and q here.

If your dataset does not work for you, please comment why. For example, the dataset is too short, exhibits change of behavior, does not look like a second-order process, etc. You might review slide 47 of Week 4, Lecture 8.

Note that you may change the project dataset in the future. This question is designed to help you start planning your the project. Please include plots of corresponding theoretical acfs and the corresponding R code.

3. An ARMA(3, 0) model is fit to the following quarterly time series:

Year	Quarter 1	Quarter 2	Quarter 3	Quarter 4
2018	3.53	1.33	1.85	0.61
2019	0.98	3.61	3.44	3.38
2020	2.91	2.12	4.62	2.93

The estimated coefficients are:

ar1	ar2	ar3	intercept
0.252	0.061	-0.202	2.637

Forecast the value for Quarter 1 of 2021. Give full explanation on how you arrived to your answer. Show calculations.

A. Less that 3.00 B. At least 3.00, but less than 3.25 C. At least 3.25, but less than 3.50
D. Atleast 3.50, but less than 3.75 E. At least 3.75.

Important: For models with AR part, the "intercept" reported in standard output of R is a misnomer. It is actually the mean of the process, so that the fitted model is

$X_t - 2.637 = 0.252(X_{t-1} - 2.637) + 0.061(X_{t-2} - 2.637) - 0.202(X_{t-3} - 2.637) + Z_t.$

4. You are given the following information about an AR(1) model with mean 0: $\rho(2) = 0.215$, $\rho(3) = -0.100$, $X_T = -0.431$. Calculate the forecasted value of X_{T+1}.

5. The five models, AR(1), ARMA(1, 1), ARMA(1, 2), ARMA(2, 3), and ARMA(4, 3) are fitted to the same time series. The models are ranked using Akaike Information Criterion (AIC): AIC $= -2\times$ log-likelihood$+2\ \times(p+q+2)$.

You are given the following information:

Model	Loglikelihood
AR(1)	-650
ARMA(1, 1)	-641
ARMA(1, 2)	-636
ARMA(2, 3)	-630
ARMA(4, 3)	-629

Determine the best model.

The following problems are for students enrolled in PSTAT 274 ONLY

G1. Suppose that in a sample of size 100, we obtain $\hat{\rho}(1) = 0.438$ and $\hat{\rho}(2) = 0.145$. Assuming that the data were generated from an MA(1) model, construct approximate 95% confidence intervals for both $\rho(1)$ and $\rho(2)$. Based on these two confidence intervals, are the data consistent with an MA(1) model with $\theta = 0.6$?

Help: (i) Bartlett's formula, was discussed in §10.3, slides 52, 55 and 56, of week 4. It gives distribution of sample acfs: For large sample size n, $\underline{\hat{\rho}}_h$ is approximately $\mathcal{N}(\underline{\rho}, n^{-1}W)$ with the elements of matrix W computed by Bartelett's formula:

$$w_{ij} = \sum_{k=1}^{\infty}\{\rho(k+i)+\rho(k-i)-2\rho(i)\rho(k)\}\times\{\rho(k+j)+\rho(k-j)-2\rho(j)\rho(k)\}$$

(ii) The acf of MA(1) model was calculated in §3.2, week 1, see slide 64.
(iii) To write confidence intervals for $\rho(1)$ and $\rho(2)$ you need to compute w_{ii} for $i = 1, 2$ from Bartlett's formula. For example, for MA(1) one gets $w_{11} = 1 - 3\rho(1)^2 + 4\rho(1)^4$. Recall that $w_{ii} = 1 + 2\rho(1)^2$ for $i > 1$ (Why?).