MSDS 413, Assignment 4
ARMA Models with Seasonality (TS4)

Introduction

For this assignment, you will use the datasets and R script (TS4.R) included in the zip file (TS4.zip), posted to the Module 4 Overview page of Canvas. You will read the data files into R and conduct the requested analyses.

The instructions for submitting your assignment follow the Procedure section below.

The following list defines the data sets and their respective variables.

Global Land and Ocean Temperature Anomalies, November (`https://www.ncdc.noaa.gov/cag/global/time-series/globe/land_ocean/1/11/1880-2020/data.csv`), Units in Degrees Celsius, Base Period is 1901-2000, Missing data indicator is -999.

- Decades by year of global land and ocean temperature anomalies.
  - Year beginning with 1880 and ending with year 2020
  - Value: Average annual temperatures deviation from base period in degrees Celsius

Your objective is to explore the time series behavior of these data sets including EDA, modeling, model diagnostics, and interpretation.

Procedure

The following steps are necessary to complete this assignment. Address each and every part and ensure that you cover all the details specified in the questions.

1. **EDA: Global Land and Ocean Temperature Anomalies** (20 points) Conduct a complete EDA on the global land and ocean temperature anomalies.
2. **Seasonal Autoregressive Moving Average (SARMA) Models** (20 points) Based on the EDA from part 1, construct a SARMA model for the global land and ocean temperature anomalies as follows.
   2.1. Your EDA should have identified a trend. Justify that this trend has been removed.
   2.2. Write the equation ($x_t = c + \phi_1 x_{t-1} + \cdots + \theta_1 z_t + \theta_2 z_{t-1} + \cdots =$) of a ARMA$(p, q)$ model then construct an ARMA(p, q) model both based on your choice of d for the temperature deviation series. Perform model checking using lag = 20. Is the model adequate? Why?
   2.3. Fit a seasonal model for the temperature series using the command (d from your EDA) **ms <- arima(y, order=c(0,0,0), seasonal=list(order=c(1,d,1)), include.mean=F)** Perform model checking including using lag=20 and adjust P and Q to improve the model if needed. Is the seasonal model adequate? Why?

2.4. Based on in-sample fitting, which model is preferred? Why?

2.5. Consider out-of-sample predictions. Use $t = 100$ as the starting forecast origin. Which model is preferred based on the out-of-sample predictions?

3. **ARMA×SARMA Models** (20 points) Continuing with the global land and ocean temperature anomalies data, construct an ARMA×SARMA model as follows:

3.1. Fit a seasonal model for the temperature series based the ACF and PACF and your choice of d using the command
ms <- Arima(y, order=c(p,d,q), seasonal=list(order=c(1,0,1)), include.mean=F)
Perform model checking including using lag=20 and adjust p, q, P, and Q to improve the model if needed. Is the seasonal model adequate? Why?

3.2. Based on in-sample fitting, which model is preferred? Why?

3.3. Consider out-of-sample predictions. Use $t = 100$ as the starting forecast origin. Which model is preferred based on the out-of-sample predictions?

3.4. Compare your SARMA model with your ARMA×SARMA model using the model diagnostics and choose which is better.

3.5. Compare your SARMA model with your ARMA×SARMA model using the model fit statistics and forecasting ability. Which is better? Do the two comparison methods agree?

4. **Report** (20 points) For the global land and ocean temperature anomalies, describe to a client or employer your best model. The report requires information from which decisions can be made or actions taken.

Deliverables

Your instructor may modify these and all the following directions. See Section Submission Directions below. The assignment deliverables, each in pdf format, are as follows:

- *Only if requested by instructor*
  - The program or script
  - Logs
  - Outputs
- Mandatory
  Data analysis write-up: no programs, logs, or just code outputs; **complete EDA and model diagnostics are expected unless otherwise instructed; I will be looking for innovative interpretations in the assignments over and above the rote adherence to assignment requirements. Only partial credit will be awarded for rote adherence to assignment requirements.**.

The data analysis must follow and use the item numbering of each assignment, i.e., use the numbers, say, 1 - 5, with the sub-lettering if used. These deliverables are provided according to the instructions in the Submission Directions section below.

Submission Directions

Title Page

Include a title page with your name and the assignment designation. Leave room for instructor comments.

File Names

The assignment write-up file shall be submitted to Canvas according to the schedule in the syllabus using the item (1) naming convention below. The naming convention is case sensitive. Use letters and numbers as given. **The file name parts have no spaces or other separator characters**. TS4Lastname.pdf (submit via Canvas)

The parts are the assignment code, TS4; your last name with only the first letter capitalized; a period, and lastly, the extension "pdf". Generically,

TS4Lastname.pdf

For example: Suppose your name is Student McStats. Your filename then is:

TS4Mcstats.pdf

The analysis write-up file must be submitted for grading. Each write-up requires a title page for instructor comments. The analysis may use either R or any other statistics package you wish, or if you use more than one package, you must use the germane tables, plots, etc., in a single report. If you use more than one package, differences and similarities should be indicated.

email

ONLY IF REQUESTED email your instructor the program (script or code), log and output as separate pdf files. The R log and output may be combined. The file names shall be as follows:

- The program or script file names
  - TS4LastnameRprog.pdf
- The log file names
  - TS4LastnameRlog.pdf