MSDS 413, Assignment 3
ARMA Models (TS3)

Introduction

For this assignment, you will use the datasets and R script (TS3.R) included in the zip file (TS3.zip), posted to the Module 3 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.

- Covid-19 data of interest (`https://covid.ourworldindata.org/data/owid-covid-data.csv`)
  - date: ISO 8601 date (YYYY-MM-DD) of the datapoint, e.g., 2020-03-21
  - iso_code: country code
  - total_cases: confirmed cases
  - population

Your objective is to explore the time series behavior of the sentiment data.

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** (20 point) Consider the data set of daily total number of Covid-19 cases confirmed after positive test. The data file is https://covid.ourworldindata.org/data/owid-covid-data.csv with column names as given above. Use the columns date and total_cases. Use your EDA from Assignment 1 to obtain and justify a stationary total_cases time series. You may need to log-transform or first difference or both.

2. **Moving Average (MA) Models** (20 points)

   2.1. Use the ACF to determine the order to fit a MA model. Justify your choice of order.

   2.2. Use the command **auto.arima**($y, d = 0$,max.p=0,stationary=TRUE), where y is your stationary total_cases time series, to find the MA order. Interpret and compare with your ACF choice. Give the model degrees of freedom (df).

   2.3. Construct a MA() model from either part 2.1. or part 2.2. Perform model checking to validate the fitted model. Interpret the diagnostics. What are the business cycles in the Covid data? What do they mean?

2.4. Obtain 1-step to 7-step ahead points with 95% interval forecasts for the total_cases data using the model you chose in part 2.3..

2.5. Forecast total_cases with the forecast origin the last observed data point using the model you chose in part 2.3.. Interpret.

3. **Autoregressive Moving Average (ARMA) Models** (20 points)

3.1. Use the ACF and the PACF to determine the order to fit an ARMA model to the Covid data. Justify your choice of order.

3.2. Use the command **auto.arima**($y, d = 0$,stationary=TRUE), where y is your stationary total_cases time series, to find the AR and MA orders. Interpret and compare with your ACF and PACF choices. Give the model degrees of freedom (df).

3.3. Construct a ARMA() model from either part 3.1. or part 3.2. Perform model checking to validate the fitted model. Interpret the diagnostics. What are the business cycles in the Covid data? What do they mean?

3.4. Obtain 1-step to 7-step ahead points with 95% interval forecasts for the total_cases data using the model you chose in part 3.3..

3.5. Forecast total_cases with the forecast origin the last observed data point using the model you chose in part 2.3.. Interpret.

4. **Report** (20 points) Choose the "best" model outcomes from the parts above. Write an executive report with information from the analysis such as forecasts from which decisions or actions can be made or 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**. TS3Lastname.pdf (submit via Canvas)

The parts are the assignment code, TS3; your last name with only the first letter capitalized; a period, and lastly, the extension "pdf". Generically,

TS3Lastname.pdf

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

TS3Mcstats.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
  - TS3LastnameRprog.pdf
- The log file names
  - TS3LastnameRlog.pdf