Assignment 2

STAT317/ECON323

Due 5pm Friday, 16 September 2022

A reminder that graphs help in interpretation and explanation and that you are expected to present them properly.

1 Question 1 – 11 Marks

For assignment 2 I have supplied a text file (rainfall.dat). This has the hourly rainfall data at Christchurch from 00:00 1 January 1992 to 23:00 31 December 1992. That is, the first hour of 1992 to the last hour of 1992.

a Get the data from the text file and put it into an R date-time series object. The date-time variable is fairly standard structure so you should be able to read it directly – though you may need to search the web to find out how. But I am OK with just importing the rainfall data and you adding date-time by hand. Whatever works.

b Since most of the time the hourly rainfall is zero (0) these zero measurements are not in the file (hence it starting 5 January, NOT 1 January). Create a complete time series with all 8,784 hours in the series. That is, add the hours with zero rainfall. Plot this series. What percentage of the hours have zero rainfall? Ideally inside R, but if you find it easier to do in a spreadsheet then import into R, that's OK, though a lot more work.

c Produce an ACF plot for this series. What does this tell you about rainfall?

d From the hourly data create a time series of daily data. Ideally inside R, but if you find it easier to do in a spreadsheet then import into R, that's OK.

e Produce an ACF plot for the daily rainfall series? Comparing it with previous ACF how similar or dissimilar are they? Explain your result?

2 Question 2 – 9 marks

For this question use the time series you selected for Assignment #1. Use the data from the years 2000-2019 only i.e. suppress COVID effects. It is recommended you use the HoltWinters option in the forecast package.

a Fit the following models to your data using the following methods:

(a) Single exponential smoothing

(b) Exponential Smoothing with trend

(c) Exponential Smoothing with trend and seasonal component

(d) The previous model but applied to a log transformed series;

For each of the models, is the time series of the residuals what you would expect for a proper time series model fit? Also from the residuals, which do you think is the best model? Explain your reasoning.

b For model 1(d) what are the values of for α, β, γ? What does this tell you about the weighting for measured recent data compared to that for estimates?