Math339 **Fall 2022**

Test01

Trun-in needs to be submitted through the coursesite link in PDF format.

1. (40 points) Table B.22 contains data from the Danish Energy Agency on Danish crude oil production.

   (a) Plot the data and comment on any features that you observe from the graph. Calculate and plot the sample ACF and variogram. Interpret these graphs.

   (b) Apply both first-order and double exponential smoothing to develop two forecasting models for crude oil production. Plot two smoothed fits on the same axes with the original data and visually compare their performance.

   (c) Apply a first difference on the raw data, plot the data and comment on any features that you observe from the graph. Generate and interpret the sample ACF and variogram.

   (d) Apply first-order exponential smoothing on the **differenced** data and use the smoothed difference to develop a forecasting model for crude oil production (i.e. $\hat{y}_{t+1} = y_t + \hat{d}_{t+1}$, $t = 1, 2, \ldots$ and $\hat{y}_1 = y_1$; $\hat{d}_{t+1}$ is the smoothed difference at time $t+1$). Compare the result to (b) where the smoothing was done on the raw (**undifferenced**) data.

2. (60 points) Table B.23 shows the weekly data of positive laboratory test results (in percentage) of influenza from the 40th week of 1997 to the 31st week of 2014.

   (a) Manage to read the data into R (or Python if choose to use) with corresponding time indices. (Note that you need to convert the information of year and week number into a valid time index) Explain how you convert the excel worksheet into an operable object in R (or Python).

   (b) Notice that these data have a number of missing values in year 1998-2002. Read the book section §1.4.3 for data imputation, evaluate the structure of data by plotting the subset of data where there is no missing values(say 2003 and on) and apply an imputation scheme that you think of as appropriate for the data. Display the imputed data in a series plot and explain your choice. (note that "xts" package provides several functions for imputation, though they might not be your choice)

   (c) Use the data from 1997–2013 to develop a multiplicative Winters-type exponential smoothing model for the imputed data. Evaluate the forecast errors to see if they significantly differ from a set of white noises (in addition to ACF, consider using Ljung-Box test which are available in both R and Python).

   (d) Use this model to make one-week-ahead forecasts for the year 2014. Plot the forecasts overlapped with the raw data and report the sum of the forecast errors. Discuss the reasonableness of the forecasts.

   (e) Repeat (c) and (d) but with an additive Winters-type model.

   (f) Compare the performance of the additive and the multiplicative model. Comment on the superiority between the two.

   (g) Visit *https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html* and download the data for the rest of year 2014 to 2021. Manage to combine the data and use the winner you pick from (e) to fit the whole set of data (1997-2020). Comment on the result.