The University of Nottingham

SCHOOL OF MATHEMATICAL SCIENCES

SPRING SEMESTER 2022-2023

MATH4022 - TIME SERIES AND FORECASTING

Time Series and Forecasting - Coursework Assignment

Answer **ALL** questions.

*Your neat, clearly-legible solutions should be submitted electronically as a pdf file via the MATH4022 Moodle page by the deadline indicated there. When typing your answers, a font **NO SMALLER** than size 11 points should be used.*

Solutions to all questions should be combined into one pdf file and submitted via the link within the 'MATH4022 - Assessed Coursework' section of the MATH4022 moodle page. Your submitted file should be named as follows:

'STUDENTNUMBER-MATH4022-CW.pdf'.

Here, STUDENTNUMBER should be replaced with your student number. For example, if my student number is 1234567, then I would save this file as

'1234567-MATH4022-CW.pdf'

You should also include your name on the first page of your submitted work.

Since this work is assessed, your submission must be entirely your own work (see the University's policy on Academic Misconduct).

Please note: the deadline is strict and submissions up to five working days late will be subject to a penalty of 5% of the maximum mark per working day. Submissions made later than five working days after the deadline will receive a mark of zero for assessment purposes.

1. The dataset `a23_nox.csv` (available from the 'MATH4022 Assessed Coursework' section of the MATH4022 moodle page) contains daily mean measurements of the nitrous oxides level (in ug/m^3), recorded at a location on the A23 Purley Way road in the London Borough of Croydon from 1st February 2017 to 30th September 2017 (inclusive).

   Once the dataset has been read into R (using the command `read.csv`), the nitrous oxides level may be set as time series data using the command:

   `ts(a23_nox$daily_mean_nox).`

   Your task is to analyse this dataset by fitting a suitable time series model to describe these data. You should write a short report to describe how you have identified and checked the fit of your chosen model. The equation of your final fitted model should be included in your report.

   Please note, there is no unique, correct, solution to this task. Credit will awarded for appropriate justification and explanation of your model choice and communication of your results.

   The R code used to produce your report should be suitably commented and included at the end of your report as a separate appendix for marking. Your report should be no longer than **SIX** A4 pages in length, including all plots and/or tables but not including the appendix that contains your R code.

   This task will marked out of 10 with marks allocated as follows:

Technical content, appropriate methods and justification of your modelling:	4 Marks
Overall presentation, writing and interpretation:	4 Marks
Correct and appropriately commented R code:	2 Marks

2. Suppose that you work as a statistical advisor to the car industry. Your task is to analyse a dataset that contains the number of new cars registered in England for each quarter (of a year) and forecast some future numbers of new cars registered.

   The dataset `eng_car_reg.csv` (available from the 'MATH4022 Assessed Coursework' section of the MATH4022 moodle page) contains quarterly numbers of new car registrations in England (in thousands), from Q1 2001 to Q3 2022. Note that one quarter = 3 months where, in a given year, Q1 = January to March, Q2 = April to June, Q3 = July to September and Q4 = October to December.

   Once the dataset has been read into R (using the command `read.csv`), the number of new cars registered (in thousands) each quarter may be set as time series data using the command:

   `ts(eng_car_reg$no_new_reg,start=c(2001,1),frequency=4).`

   Using this dataset, your task is to produce a time series model to forecast the number of new cars registered for Q4 of 2022 and Q1-Q3 of 2023.

   You should produce a short report that would be suitable for your peer statisticians/data scientists to understand. Your report should contain:

   (i) The main aim(s) of your work.

   (ii) A clear and succinct description of how you analysed the data and arrived at your final model.

   (iii) Information on how you checked model fit.

(iv) Information on the forecasted numbers of new cars registered for Q4 of 2022 and also for Q1, Q2 and Q3 of 2023, together with associated uncertainties.

The first page of your report should be an 'Executive Summary' in which you convey the main findings of your work. The executive summary must be no longer than one page, consisting of text and at most one plot (with font size no smaller than 11 points). The executive summary must be concise and suitable for a broad set of non-statistician readers to understand, including the head of a car company sales division.

The R code that you used for your work should be suitably commented and included as an appendix to your report for marking.

Your report should be no longer than **12** A4 pages in length, including all figures and tables but not including the R code appendix or one-page executive summary. When typing your work, a font size no smaller than 11 points should be used. Please note: this does not mean that a completely correct solution would need to be a 12 page report.

Throughout your report you should aim to convey important details of your work in a way which is easy to follow but not excessively long and credit will be given for this. When producing your report, think about someone reading it through and try to make it easy for them to understand your work. Try to be clear in your writing, without too much repetition, and avoid long items of numerical output.

This task will be marked out of 30, with marks allocated as follows:

Technical content, appropriate methods and justification of your modelling:	10 Marks
Overall presentation, writing and interpretation of results:	10 Marks
Executive summary:	5 Marks
Correct and appropriately commented R code:	5 Marks