F70TS TIME SERIES
Coursework 2023/24

INSTRUCTIONS

Please read the instructions carefully before you commence doing the coursework.

What to submit

The coursework must be done individually - for all parts - the coding and the writing up of your report. Your submission consists of two files:

- A Word document (`.doc` or `.docx`) or portable document format (`.pdf`), which is submitted via TurnItIn in Canvas, containing your report. It contains the answers to the 6 questions at the end of this document.

  The **report must be at most 4 sides-of-a-page in length**. For example, this current document is 4-sides-of-a-page long. Anything more will not be marked.

- A text file (`.txt`) containing the *R* code you wrote and ran to generate the plots and values in your report.

Where to submit

Submit your coursework in the F70TS Canvas site.

When to submit

You must submit your coursework before 15:30 (local time) on Friday 15 March 2024.

If you submit after the deadline

If you submit your assignment after the deadline, then

- If you submit up to 5 working days late, a fixed deduction of 30% is made to your coursework mark.
- If you submit more than 5 working days late, then your coursework will not be marked and you will receive a mark of 'no grade' for the coursework. The only exception is if you have a Mitigating Circumstances application accepted.

Marks

This project has a total of 20 marks available and will count a maximum of 20% towards your final grade.

Feedback

You will receive your coursework back in Canvas within 15 working days of the due date. A rubric attached to your submission will show where marks were awarded.

Academic misconduct, such as plagiarism and collusion

The 'Academic Integrity' module, which is on the F70TS Canvas site, explains the various forms and types of academic misconduct. If you have not already completed the 'Academic Integrity' module in another course, during this academic year, then you must complete it before submitting your coursework.

You must complete and submit the 'Declaration of Authorship' form in Canvas (found in the same module as where you find this file on Canvas).

The University takes plagiarism and collusion very seriously. It is your responsibility to ensure that you understand what you are and are not allowed to do.

Using *R*

The following packages much be installed in your *R* software:

```
install.packages("stats")
install.packages("Rcpp")
install.packages("ggplot2")
install.packages("forecast")
```

When you begin a session in *R*, you must load these packages:

```
library("Rcpp")
library("ggplot2")
library("stats")
library("forecast")
```

You must also tell *R* where to look for the data file `DataN20.csv` and then read it in to *R*, e.g. if you downloaded the data file `DataN20.csv` from the coursework Canvas module and saved it in the folder `C:\Users\abc123\coursework\`, then type in *R*:

```
path <- file.path("C:", "Users", "abc123", "coursework", fsep="\\")
setwd(path)
dataset.raw <- read.csv("DataN20.csv")
dataset.ts <- ts(data = dataset.raw[,2], start = c(2001, 1), frequency=12)
```

R commands which are not allowed - `auto.arima`

You must not use the *R* command `auto.arima` in this project. This is an explicit instruction. Additionally, it will not give you the required answers!

Forecasting levels of atmospheric nitrous oxide

Gases that trap heat in the atmosphere are called greenhouse gases. Examples are carbon dioxide, methane, nitrous oxide (N_2O) and fluorinated gases. Greenhouse gases are the most significant driver of observed climate change since the mid-20th century. Larger emissions of greenhouse gases lead to higher concentrations in the atmosphere. In turn, these act as a planetary blanket in the atmosphere and warm the climate, leading to many other changes around the world - in the atmosphere, on land and in the oceans.

It is important to monitor and forecast the levels of nitrous oxide in the atmosphere due to its prevalence in human-driven activities and its impact on the climate. Examples of when nitrous oxide is emitted are in the use of nitrogen fertilisers in agriculture, as a byproduct of fuel combustion and during the treatment of wastewater. Each N_2O molecule spends, on average, about 110 years in the atmosphere before moving to a different part of the ocean–atmosphere–land system. Additionally, an N_2O molecule traps significantly more heat than each CO_2 molecule (about 275 times more heat).

You have been asked to fit a time series model to observations of the concentration of nitrous oxide, and then forecast the likely path of the concentration over the next 3 years.

You can find the data for the task in the file **DataN2O.csv**. The data gives the globally-averaged concentration of nitrous oxide, measured in parts per billion (ppb), at marine surface sites. 1 ppb indicates that one out of every billion molecules in an air sample is N_2O. The measurements were done monthly, starting from January 2001.

Based on the data $\{x_1, \ldots, x_N\}$, answer the following questions

1. Generate a time series plot of:

- The data, $\{x_m\}$, against the observation times $\{m\}$.
- The first differences of the data, $\{x_m - x_{m-1}\}$, against the observation time m. The R command `diff` can be used to create the time series of first differences.

The plots should have labels on the x- and y-axis and, when displayed in your report, they must have a caption.

Display on the plots in your report. State for both plots if they suggest stationarity, supporting your statements with observations from the plots.

[4 marks]

2. Plot the autocorrelation function (ACF) and partial autocorrelation function (PACF), up to lag 20 months, of:

- The data, $\{x_m\}$, and
- The first differences of the data, $\{x_m - x_{m-1}\}$.

The plots should have labels on the x- and y-axis and, when displayed in your report, they must have a caption.

Display and comment on the plots in your report. The ACF and PACF should be considered together, for each data set. [4 marks]

3. Based on your observations above, state in your report a suitable ARIMA(p, d, q) model to fit to the original data, where $d \in \{0, 1\}$. [1 mark]

4. In R and using the R command `Arima`, fit the chosen ARIMA(p, d, q) model to the data.

Let $\{X_t\}$ denote the time series model from which the data is assumed is generated and let B denote the backshift operator.

By substituting the unknown parameters with values, write down the equation of the fitted model in your report in the form

$$(1 - \phi_1 B - \cdots - \phi_p B^p)\,(1 - B)^d\,(X_t - \mu) = (1 + \theta_1 B + \cdots + \theta_q B^q)\,\varepsilon_t,$$

where $\{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2)$ and μ is the estimated mean of $(1 - B)^d\,X_t$ generated by the `Arima` function. Give the parameter values to 3 decimal places.

Additionally, state the AIC value generated by the R command `Arima` for your model.

[4 marks]

5. Forecast the N_2O concentration for the next three years beyond the original data. Generate a plot of the forecast, which shows (i) the original data; and (ii) its forecasted mean with a 95% forecast interval around it. The plot should have labels on the x- and y-axis and, when displayed in your report, it must have a caption.

In your report, display the plot and comment on its implications for the concentration of N_2O in the future.

[3 marks]

6. A colleague looks at your plot of the original data set and suggests that fitting a straight line to the data, by using the 'least squares' approach (i.e. linear regression), would be far simpler.

State four (and no more) advantages of ARIMA models over linear regression, briefly explaining why they are each an advantage. Use bullet-points to structure your answer, with each bullet-point corresponding to an advantage.

[4 marks]

[Total 20 marks]