F70TS 2020-21: Assessed Project

In this project we will use the R programming language to simulate, fit and forecast time series. This project has a total of 30 marks and will count towards 30% of your final grade.

Part I

A time series is given by the formula:

$$\begin{aligned} X_t = &- 0.022X_{t-1} - 0.405X_{t-2} + 0.493X_{t-3} \\ &+ 0.550\varepsilon_{t-1} + 0.276\varepsilon_{t-2} + 0.490\varepsilon_{t-3} + 0.682\varepsilon_{t-4} + 0.222\varepsilon_{t-5} - 0.232\varepsilon_{t-6} + \varepsilon_t \end{aligned} \quad (1)$$

where for all t, the innovations ϵ_t are distributed according to the standard normal distribution, that is, $\varepsilon_t \sim N(0, 1)$.

1. Generate a realisation from this time series in R, of length 1000. Plot the realisation and also its autocorrelation and partial autocorrelation functions.

   [3 Marks]

2. The solutions to equations of the form:

   $$a_0 + a_1 x + a_2 x^2 + \ldots a_n x^n = 0$$

   can be found in R using `polyroot(a)`, where `a` is a vector of the polynomial coefficients in increasing order (that is, $a = (a_0, a_1, \ldots, a_n)$). For example, `polyroot(c(1,2,3))` will find the solution of $1 + 2x + 3x^2 = 0$.

   Is the time series given by equation (1) causal stationary? Is it invertible? The R command `Mod(z)` can be used to calculate $|z|$, the modulus of a complex number.

   [3 Marks]

3. The file `data_1.txt` contains a realisation of a time series.

   a) It is believed this time series is generated from a MA(3) model with normal innovations, that is, $\varepsilon_t \sim N(0, \sigma_\epsilon^2)$ for some $\sigma_\epsilon^2 > 0$. State the formula of this time series model. Fit this model to this observed time series. Store the estimated coefficients you obtain in the R variable `coef_1` and the estimated innovation variance in the variable `sigma_sq_eps_1` in R. State the values of the estimated coefficients that you obtained, including the estimate of σ_ϵ^2.

   [3 Marks]

b) Simulate the next 1000 values of the time series using the model fitted in (a). This simulation should be continued from the end of the data supplied to you in `data_1.txt`, starting with $X_{1001}, X_{1002} \ldots$ etc. Plot the values of this realisation and the original data and comment on the adequacy of the model that you fitted to the data in (a). (Hint: the R command `simulate` may be useful)

[3 Marks]

[Total 12 Marks]

Part II

The data used in this question can be found in the file `data_2.txt`. Fit an ARIMA(p, d, q) model with $N(0, \sigma_\epsilon^2)$ innovations to this observed time series:

1. Select the value of d. Justify your answer.

   [3 Marks]

2. It is known that the values of p and q are less than 3. Select the values of p and q. Justify your answer.

   [3 Marks]

3. Fit the ARIMA(p, d, q) model with your selected values of d, p and q from Questions 1 and 2 of Part II. State the values of the estimates of the coefficients and the variance of the innovation.

   [3 Marks]

4. Generate a 95% forecast interval for the next 500 values of the time series. Plot the forecast intervals along with the data. Briefly comment on the forecast intervals.

   [3 Marks]

[Total 12 Marks]

[END OF PROJECT QUESTIONS]

Your findings should be presented in the form of a report, which should:

- contain the answers to the each of the questions, in order, with the part and question numbers labelled;
- include detail of your mathematical calculations and the statistical methods used so that your results could be reproduced by another statistician;
- include clearly labelled and correctly referenced tables and diagrams, as appropriate;
- include the R code you used in an appendix (you do not need to explain individual R commands but some comments should be included to indicate the purpose of each section of code). **Do not include figures, references, or mathematical developments in your appendix, only R code with comments.**

- **maximum page limit of nine (9) pages (11-point font, A4 size)**. The R code in the appendix does not count towards this page limit.
- include citation and referencing for any material (books, papers, websites etc) used. When possible, use reliable sources, produced by respected and well-known authors, published by recognised publishers and associated with well established government, academic, or educational institutions. Note that some webpages, YouTube videos, blog posts, and Wikipedia pages might include errors.

A total of **6 Marks** is available for these aspects of your report. This will be marked according to the rubric given in the Appendix.

Each question in this project is allocated 3 marks:

- 3 marks will be awarded for an answer that shows a clear understanding of the mathematical and statistical theory, have clear and efficient R code, and come to the correct conclusions.
- 2 marks will be awarded for an answer with minor errors. For example, the correct formulae were stated, but then a wrong number was incorrectly substituted, or a minus sign was missed on the next line. Such answers will show an understanding of the theory and R coding, but there may be flaws in the calculations or code.
- 1 mark will be awarded for an answer with major errors. Such answers will show major flaws in understanding of the theory or the R programming language.
- 0 marks will be awarded for an answer which neither demonstrates an understanding of the theory, and the R code is incorrect/will not run.

Notes

- This assignment counts for 30% of the course assessment.
- You may have face-to-face discussions with me or your colleagues, but your report must be your own work. **Plagiarism** is a serious academic offence and carries a range of penalties, some very serious. Copying a friend's report or code, or copying text into your report from another source (such as a book or website) without citing and referencing that source, is plagiarism. **Collusion** is also a serious academic offence. You must not share a copy of your report (as a hard copy or in electronic form) or your computer code with anyone else. Penalties for plagiarism or collusion can include voiding of your mark for the course. See `https://www.hw.ac.uk/students/studies/examinations/plagiarism.htm` for more details.
- **Your report should be submitted through Turnitin by 15:00 GMT on Monday 29 March 2021. A link to the submission page is available through the 'Assessment' section of the course Vision page. Please use the submission link appropriate for the campus where you are studying (Edinburgh or Malaysia). For late submissions, 30% will be deducted for work submitted at most 5 days late. Submissions that are more than 5 days late will receive 0 marks.**

Appendix: Rubric for marking of the report

The six marks available for the exposition of your report will be awarded according to the scale below:

0–1 Marks will be awarded for	• Lack of clear and logical structure • Conclusions missing or not suitable for a non-statistician • Statistical calculations and methodology not clearly set out for the reader • Tables and figures unclear, badly labelled or not correctly referred to • R code not included, or no comments included in it • Sources used not clearly referenced
2–4 Marks will be awarded for	• Clear and logical structure • Conclusions generally suitable for a non-statistician • Statistical calculations and methodology generally set out clearly for the reader • Tables and figures often clear and correctly referred to • R code included with some comments • Sources used clearly referenced
5–6 Marks will be awarded for	• Clear and logical structure • Conclusions suitable for a non-statistician • Statistical calculations and methodology set out clearly for the reader • Tables and figures clear, correctly referred to and easy to interpret • R code included with comments • Sources used clearly and correctly referenced

You will receive feedback during the week beginning Monday 19 April 2021.