

2022/2023 MSCI 523 Forecasting Coursework - Part 1

Coursework Information & Submission

This is the first of two assignments. This first GROUP assignment is weighted 40% and will require you as a group to explore a dataset of time series. Each person will be assigned a unique time series. You can form a group of 5 students yourself and will then analyse 5 time series in total together as a team.

Completing this assignment will prepare you for the 2nd individual assignment, weighted 60%, which will be to forecast the same time series you are assigned now with multiple forecasting algorithms.

Coursework deadline is **20.03.2022, 10:00am**.

Standard departmental penalties will apply for late work unless you have been given an extension for exceptional reasons from the course administrator. All submissions will be checked by the plagiarism software. Coursework must be submitted online on Moodle. Submit your report PLUS all R scripts in the appendix in Moodle.

Assignment: Data Exploration of a real-world time series

Your task is to **explore and critically discuss the time series pattern(s) of a real-world time series**. Document your findings comprehensively in a technical report, making adequate use of (readable and correctly labelled) graphs which you also critically discuss to support your arguments. Base your justification on evidence and document your iterative data exploration process, possibly transforming the time series and analysing the resulting patterns throughout.

20 % of points – Explore time series graphically & verbally.

Explore the regular components and the irregular components of the time series making good use of graphs, plots, and **descriptive summary statistics**. Critically **discuss the patterns** you observe verbally, both regular (level, trend, season, etc.) and irregular patterns (outliers, breaks, etc.) and conclude what patterns are observed.

20% of points – Statistical Tests of time series

Explore the data using statistical tests. Complement your analysis with statistical tests to support your visual analysis, **interpret their outputs, and discuss potential discrepancies with your visual analysis**. Note that important tests include **stationarity (e.g. the ADF test), regular time series patterns (e.g. seasonality, trend, etc) and irregular time series patterns (outliers, structural breaks)**.

20 % of points – Explore Temporal Aggregation of Time Series

Consider transforming the time series to exemplify individual patterns, both by **removing patterns or transforming the time series into an aggregate form of lower frequency (e.g. quarterly, monthly, weekly buckets instead of daily)**, and then **exploring these graphically, statistically and commenting on the identified patterns**.

10% of points – ACF Analysis of time series

Explore the data using ACF & PACF. Plot the **ACF and / or PACF graphs** of the time series as suitable. Also consider to **iteratively transform the time series** depending on the patterns to consider non-stationarity, trends and / or seasonality and **demonstrate the effect of transformations on ACF/PACF graphs** with correct interpretation.

10% of points – Summarise across all time series

Conclude by **summarising the patterns across all individual time series** in your group, **comment on commonalities of time series patterns as well as differences**, are they similar or dissimilar, and **what this implies to forecasting many time series of this sort automatically**.

10% of points – Conclusions

Conclude by **recommending one (or multiple) suitable algorithm(s) and forecasting model form(s)**, and **critically discuss your choice weighting the different options**. As time series patterns are not always clear, there often are multiple suitable forecasting model for each time series. Please **recommend all that are suitable**.

10% of points - General report writing skills

General report writing skills include a critical discussion of findings, thoroughness of documentation, clarity of arguments, structure of the report, readability of the report (i.e. lack of spelling and grammatical mistakes etc.) in marking each section. Please see next page for some more technical considerations on report writing.

SUM 100%

Although you are free to use any external software, we highly recommend the use of R (e.g. as python will not the same amount of statistical tests and thus limit your ability to achieve full points). Report the software you used.

Please also consider the general recommendations on writing a technical report on the next page!

Forming a Group

You are invited to form a group of 5 students by yourself – please approach other students in the class directly. If you cannot form a complete group of 5, or you cannot find a group to join, then you will be allocated to a group by us. Please send the names and library card numbers of all group members to me by email at s.crone@lancaster.ac.uk with “MSCI523 GROUP” as a reference until **Wednesday 20.02.2022 15:00**. After that, all students without a group will be allocated randomly to a group and groups can no longer be changed.

General suggestions on writing a report

The coursework requires you to document your analysis and critically discuss your chosen experimental design, modelling approaches and the results in a technical report. This technical report should be written as if tailored to an Analytics specialist (e.g. who has an MSc from Lancaster University and has taken the MSCI523 course, and who wants to evaluate your results AND your decision making process to determine your skills in modelling and whether you have missed anything). This means that you are not required to write a general description (i.e. a statistical test is, the ACF function is, Exponential Smoothing is ...) as an Analytics expert would be aware of this! Consequently, the report should document the process of modelling, and allow an understanding of your choices and a replication of your experiments.

The report should contain an introduction and a summary with conclusions on your findings, numbered headings, list of figures and tables and an executive summary (tailored to senior management) indicating the most relevant findings. The report should display a logical and concise structure, be generally “readable” and support your argument using plots of time series, forecasts and /or accuracy. For formatting, use single spacing, format normal text in times new roman font size 11 or 12, text in tables, figure and table headings in font size 9 or 10 as long as they are readable when printed, and leave 2cm of margin left and right. Make adequate use of graphs to show time series, model fit / predictions and residuals to support your arguments (graphs must be completely readable and with all axis labels when printed on A4), as well as tables to compare results.

The page limit for the main part of the report is 2 pages fixed plus 8 pages per each time series, i.e. 42 pages (2+5*8) for a group of 5 or 34 for a group of 4 students. Pages count only towards the main text incl. graphs and tables, but not for the cover sheet, executive summary, contents sheet or appendices). Reports of excessive length will be penalised by deducting 10 marks (i.e. 10% of 100) but only if they are including un-necessary material.

Include any technical details and hardcopies that support your arguments further in a set of appendices. Some analysis provides a lot of evidence (e.g. from the complete information printout from a single ADF-statistical tests you place only the conclusion of significance / insignificance with the probability in the main text and place all additional information including printout of the test output in the appendix). All parts of the text supported by an appendix must cross-reference directly to the relevant part, i.e. the section in the appendix or the numbered figure in the appendix must be referenced directly at the corresponding place in the main text, else it is not taken into consideration to award points. You must ensure the main text is readable and that your argument is coherent without needing to consult the appendices (i.e. do not move too much in the appendix as the main text needs to provide your main arguments without having to flip back and forth to the appendix). Your r-script MUST be included in the appendix of the word document of the same file – note again that the appendix does not count towards your page limit!

Non-disclosure clause: these datasets and the coursework task is subject to copyright © by Sven Crone, all rights reserved. In downloading the documents and submitting the assignment for assessment the copyright agreement is deemed accepted. Any publication of the dataset, the coursework task, or its solution (e.g. on a coursework website or a social network site), or a part thereof, will be considered a violation of copyright. The person breaking the copyright may be held liable for damages by international law suit. Furthermore, the publication will count the assignment as a plagiarism - even in retrospect after receiving the MSc degree - leading to a mark of zero, with the usual right to appeal to university court in official hearing,

Contact details:

Questions regarding the coursework
Sven F. Crone
Room A53a
s.crone@lancaster.ac.uk

Questions regarding R and workshops:
Carlos Rodriguez
c.rodriguezcalderson@lancaster.ac.uk

If you have any questions, please don't hesitate to contact us! However, please make sure you arrange a meeting first via email to avoid people queuing and being disappointed. Also consider in your enquiries that I cannot always react within a few hours, so don't leave questions to the last minute ... start early!

Best of luck!
Sven & Carlos