TSAC: Individual Project

Due Date: March 29th, 2025.
Projects turned in after 29/03/2025 at 5.00pm will receive at most 1/2 credit.

The project is a written exposition that thoroughly describes the complete analysis of a data set. The project should be submitted as a notebook on Google colab.

1 How to Choose Your Data

Here are some guidelines for choosing a data set:

- Use the internet. There are a great deal on online sources for data of all type. Do not use data available from course or book packages.
- The responses Y_t should be continuous in nature (not binary or strikingly discrete). Also, make sure you completely understand the sampling frequency; e.g., are the data collected every day? every week? ...
- Choose a data set in an area you are interested in!
- Don't choose a data set that is very small (e.g., let's stay away from series where $n < 50$). Ideally, we want $n > 75$ or so.

2 Outline of The Written Project

in this order:

- Title page and abstract. You must prepare a title page with an appropriate title and abstract. An abstract is a very high-level written summary of the entire project. Main points and findings only. (a paragraph)
- Introduction. This part introduces the reader to the data set and the area to which it pertains. Basically, introduce the reader to the problem and why it is worth investigating. This should be written at a very basic level (i.e.,no mathematics or notation). (Half a page)
- Model specification. This is the most important part of the paper and will be the longest in length. In this section, you want to describe, in clear detail, the data analysis used to specify your candidate models. (as long as needed)
- Fitting and Diagnostics. This part of the project should describe the model fitting and diagnostics techniques you used, with the goal of identifying a "final" model for forecasting. Identify also what possible deficiencies your final model has. Remember, no model is perfect.
- Forecasting. This section should describe the techniques you used to forecast future observations. It might be a good idea to "withhold" some of the data from your series towards the end of it so that you can compare your forecasts to the actual values of the process. For example, suppose that you have a series of length $n = 100$. Perform the specification, fit, and diagnostics on the first 95 observations and withhold the last 5. Then, when you forecast, you can compare your first 5 forecasts to the last 5 observations.
- Discussion. Here you want to offer a summary of what you did in the project and draw your main conclusions. Also, discuss here other issues related to the data analysis. for example, What were the main problems you encountered?

3 General Advice:

- This is a small project! So don't over do it. Keep the size of the report reasonable with a maximum of ten pages (there should more graphs and tables than text).
- Keep everything double spaced.
- Break your report into sections. Each section should have a title. Use subsections (with titles) if necessary.
- Integrate your graphics and output into the written text as you see fit. For example, if you want to show me the time series itself, embed it into the written work.

4 Grading Scale:

Your report will be graded out of a total of 10 points, based on Writing, Analysis, and Context. For example:

- Writing (out of 2 points): How organized, clearly written, comprehensible, and grammatically correct is the report? Would the client reading this report be confident that it was written by an educated, well-trained statistical scientist?
- Analysis (out of 6 points): Were the chosen models, graphs, and data analyses appropriate for the problem? Were the analyses carried out correctly? Were your statistical conclusions about the data set sensible and clearly justified by numerical or graphical evidence?
- Context (out of 2 points): Were the questions answered in terms of the variables of the data set? Have you attempted to frame your conclusions and interpretations in a subject-matter context rather than treating the data as simply a meaningless set of numbers? Have you provided some background information about the data set and why it is of interest?