STA 137—Applied Time Series Analysis — Project Outline

Data Processing

- *"Data acquisition":* Please find your own time series data. One option to find appropriate data is to go to `google.com/trends` and to search for a phrase of interest to you. Google allows you to store the resulting search data in a .csv file that can then be loaded into R. Make sure to select a data set that has clear trend and seasonal components so that you can perform a reasonable time series analysis.
- *Data visualization:* In all the following steps (also in the Data Analysis part below), make sure to provide appropriate plots in order to motivate your choices. This is a necessary non-mathematical, common sense approach to the data analysis.
- *Data preparation:* The data may need some initial cleaning. Are there missing values? Are there outliers? And so on. Before any time series analysis can start, these issues may need addressing.
- *Define the goal:* Make sure there is a clear aim for your time series analysis. Writing "I would like to forecast the next value" is not ideal. Have a good motivation as to why you chose your data and why it should be useful to look at prediction of future values. What is the information you would like to extract?

Data Analysis

- *Initial data transformations:* Check if it is necessary to stabilize the variance of the data or if transformations to symmetry are necessary. Then start modeling with the classical decomposition of a time series into trend plus seasonality plus stationary errors.
- *Analyzing the "smooth" component:* Perform a trend and seasonality analysis, choosing from the methods given in Sections 1.3 and 1.4 of the Lecture Notes. After this step, the resulting residuals should pass as stationary time series.
- *Analyzing the residuals:* Check the residuals for whiteness, remaining trends and normality as outlined in Section 1.5 of the Lecture Notes. Hopefully your residuals display some stationary time series features.
- *Analyzing the "rough" component:* Fit stationary ARMA models to the residuals obtained after the analysis of the smooth component, as outlined in Chapter 3 (see Section 3.6 for a summary) of the Lecture Notes. Make sure to provide ACF and PACF plots for support. Check if the residuals conform to white noise. If not, there is still dependence left in the data that could be utilized for improved modeling.
- *Predict future values:* The prediction of the time series is given by the prediction of the smooth component plus the prediction of the rough component. For the smooth part, you need to make sure to be able to predict trend and seasonality. For the rough part, use one of the algorithms in Section 3.5 of the Lecture Notes.

Write-Up

- *Style:* The write-up should take the form of a scientific paper. The following parts should be included: (1) Title, author(s) and abstract; (2) Introduction; (3) Data description; (4) Data analysis; (5) Discussion; (6) Conclusions; (7) References; (8) Appendix.
- *Title, author(s) and abstract:* Find a meaningful title and provide your names. The abstract should contain a concise summary of what is to come in the paper.
- *Introduction:* This part should broadly answer the questions: What is the problem considered, why is it interesting? What is the proposed solution (and how does it relate to existing literature)? How is the rest of the write-up organized?
- *Data description:* A detailed introduction to the time series data should be given. What is it that you are looking at? Where did it come from? What are the interesting features? Give summary statistics and plots to make your points.
- *Data analysis:* This section should include the remaining items from the Data Processing and Data Analysis sections above.
- *Discussion:* Explain the findings of your data analysis. Try to be critical of what you have done and make sure you don't hush over potential shortcomings of your analysis. It is better to be aware of something that is not ideal than to try covering it up. (Importantly, this is not being viewed negatively in the evaluation of your time series analysis. It can easily happen that you face an issue you cannot deal with based on the methods you learned in STA 137. When this happens try to use your best judgement to proceed in a reasonable way.)
- *Conclusions:* Give a final verdict of your analysis.
- *References:* Name all resources you have used. Do not copy from the web and existing papers. You can and should use other resources, but they have to be clearly identified. Line-by-line copying from existing contributions will be considered plagiarism.
- *Appendix:* Put all codes and additional supporting calculations here.
- *The exposition should be clear and easy to follow. Please check spelling and grammar carefully.*

Extra Credit: Spectral Analysis

- For those of you, who would like to add to your overall score in STA 137, there is the option of including a spectral analysis of your data into the write-up. Since this topic is considered only in the final two weeks of the quarter, it is not a required part of the project.
- The spectral analysis component can add up to 25% to the maximum project score.