

Homework 5
STAT 5511
Charles R. Doss
Fall 2024
Assigned: Mon, Nov 4
Due: Weds, Nov 15

The usual formatting rules:

- Your homework (HW) should be formatted to be easily readable by the grader.
- You may use knitr or Sweave in general to produce the code portions of the HW. However, the output from knitr/Sweave that you include should *be only what is necessary to answer the question*, rather than just any automatic output that R produces. (You may thus need to avoid using default R functions if they output too much unnecessary material, and/or should make use of `invisible()` or `capture.output()`.
 - For example: for output from regression, the main things we would want to see are the estimates for each coefficient (with appropriate labels of course) together with the computed OLS/linear regression standard errors and p-values. If other output is not needed to answer the question, it should be suppressed!
- Code snippets that directly answer the questions can be included in your main homework document; ideally these should be preceded by comments or text at least explaining what question they are answering. Extra code can be placed in an appendix.
- All plots produced in R should have appropriate labels on the axes as well as titles. Any plot should have explanation of what is being plotted given clearly in the accompanying text.
- Plots and figures should be *appropriately sized*, meaning they should not be too large, so that the page length is not too long. (The arguments `fig.height` and `fig.width` to knitr chunks can achieve this.)
- **Directions for “by-hand” problems:** In general, credit is given for (correct) shown work, not for final answers; so show **all** work for each problem and explain your answer fully.

1. Shumway and Stoffer (4th ed.), question 3.11,

For the next two questions, download the file `HW5dat.rsav` from Canvas and `load(`HW5dat.rsav`)`. It has two time series objects, `dat1` and `dat2`, which you will analyze in the next two questions.

The analysis for each dataset should begin on a new page and should have as label the name of the dataset (`dat1` or `dat2`). Your job is to fit the best $\text{ARIMA}(p, d, q)$ model to each dataset that you can. Your output should be in the following format. (Points will be deducted if it is not.)

- The analysis for each question/dataset should begin on a new page and should have as label the name of the dataset.
- On the first page you should state on which page each question begins. [Note: One simple way to automate this in \LaTeX is to use `\section{}` to start each dataset and then include a `\tableofcontents`. You could also use `\label{}` and `\pageref{}` commands.]
- On the first page of output for each problem, you should first have a summary (labeled “Summary”) that provides the model chosen, parameter estimates, standard errors, and p-values in that model. Specify explicitly if you exclude a constant term. For example, “I chose an $\text{ARIMA}(1, 2, 3)$ model including a constant/intercept term. The parameter estimates were ...”. If you believe the data cannot distinguish between two (or more) models you should describe both (all) of them in this manner here.
- After the summary, should be an explanation (labeled “Explanation”). Provide a clear explanation of why you selected the model you selected. Refer to the output of your analysis, which will be below. The model selection and diagnostic techniques we have discussed in class can be discussed here. You do not need to (and should not) provide an exhaustive list of all possible models, but should rather provide explanation for which models were reasonable contenders (and why), and which model (or models) were the best out of those contenders (and why).
- After the explanation is the “Output” you refer to in your explanation. (The output may be plots or output from various commands.) All of it should be clearly formatted, and labeled or described. You do not need to provide exhaustive output from every command you have run, but you should include enough to justify all the arguments you make in your summary.

Finally, *please refer to the original/raw (untransformed) time series as X_t in your descriptions and as xx in your code.*

2. Analyze `dat1`.
3. Analyze `dat2`.