

STAT 5511 Homework 7 (Fall 2024)

Charles R. Doss

Assigned: Weds December 4

Due: Weds December 11

General formatting guidelines:

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.

Instructions: For this homework, you will analyze two data sets. Find the file `hw7dat.rsav` on the course webpage. Load it into R by running `load("path/hw7dat.rsav")` where ‘path/hw7dat.rsav’ is replaced by the full path on your hard drive to the file `hw7dat.rsav` (the syntax for which is operating system dependent). The file contains multiple data objects: `dat1, dat2`. Each of the two datasets is a separate homework question. The analysis for each dataset should begin on a new page and should have as label the name of the dataset (`dat1, dat2`). Your tasks are different for the three datasets and they are as follows.

1. For `dat1`: `dat1` has two columns, `yy` and `xx`, which are both time series. Your job is to regress `yy` on `xx` using the (“transfer modelling”) methodology we discussed in class.
2. For `dat2` your job is to use the local level type models that were demonstrated in class, using the `d1m` R package. The dataset has 4 columns. The data is related to wastewater measurements about COVID-19 from the Twin Cities wastewater treatment plant, used for monitoring the disease in the Twin Cities area. The first column may be ignored. The second column is the date of each measurement. The third and fourth columns are the measurements.
 - (a) Plot the data and the local level filter and the local level smoother like we did in class for just the third column of data. That is there is a single (one dimensional) observation equation.
 - (b) Explain the differences between the smoother output and the filter output.
 - (c) Plot the data and the local level filter and the local level smoother like we did in class based on both the third and fourth columns of the data. That is there is a bivariate (two dimensional) observation equation and two observations at each time point.
 - (d) Explain the differences between the output based on just one of the series and the smoother/filter output based on both.

Presentation/formatting rules: for question1, your output should be in the following format. Points will be deducted if it is not.

- On the first page of the assignment 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, “For the series $Y_t = X_t^{1/2}$, I chose an $\text{SARIMA}(1, 2, 3) \times (4, 5, 6)_7$ model, including 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. In the case of a regression model, you should explain the full model, meaning which lags of which variables are included in the regression model as well as what the ARMA model of the errors is.
- 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, *in Question 1 the two series are named xx and yy and you should not rename them.*