Homework 2
UMN STAT 5511 (Fall 2024)
Charles R. Doss
Assigned: Mon, Sept 30
Due: Weds, Oct 9

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.

Questions:

1. (Prediction using the cross-correlation function) Assume that $Y_t = aX_{t-\ell} + W_t$ for some number a. The series X_t leads Y_t if $\ell > 0$ and is said to lag Y_t if $\ell < 0$. Assume that $E(X_t) = E(Y_t) = 0$, that $\{X_t\}$ is stationary and that $W_t \sim \mathrm{WN}(0, \sigma^2)$ is uncorrelated with the whole series X_t. Let γ_x denote the autocovariance function of $\{X_t\}$.
   (a) Is Y_t stationary?
   (b) Compute the cross covariance function between Y_t and X_s, for any s and t. (Your answer will depend on γ_x, the autocovariance function of X_t.)
   (c) Compute the cross correlation function between Y_t and X_s, for any s and t. (Your answer will depend on γ_x, the autocovariance function of X_t.)
2. Question 2.3, Shumway and Stoffer, 4th edition (Note: The question is somewhat different than in previous editions).
3. Question 2.10, Shumway and Stoffer. For (f)(iii), you can do both analysis of the residuals as you would in a non-time series context (e.g., a QQ-plot) and analysis of the correlation of the residuals (using the ACF).
4. Consider the setup of the previous question (Question 2.10, Shumway and Stoffer), and let us focus on just the oil series. One model we might consider for the (untransformed) oil series is the random walk with drift model, $X_t = \delta_1 + X_{t-1} + W_t$ where $W_t \sim \mathrm{WN}(0, \sigma^2)$. If we let $X_0 = \delta_0$ be a constant "intercept" term, then we have checked in class that we can write $X_t = \delta_0 + \delta_1 t + \sum_{s=1}^{t} W_s$. The mean of X_t is thus $\delta_0 + \delta_1 t$. We might be interested in estimating this (linear) regression function.
   (a) Use `lm()` to regress the (untransformed) oil series on time. Print the `summary()` of the results and plot the data with the regression line. Comment briefly on the statistical significance of the coefficients.
   (b) Compute the F-statistic for testing $H_0 : \delta_1 = 0$ against $H_A : \delta_1 \neq 0$ (i.e., for testing whether there is a drift) and the corresponding p-value; you can do this by using `lm()` and `summary()`.

(c) Now we return to the random walk with drift model and the results of 4b. We want to assess whether the p-values that we computed actually mean anything. We will run a simulation study to assess this, as follows.

Simulate a random walk with no drift X_t, for $t = 1, \ldots, 545$. (You may want to use the `cumsum` function.) Assume $W_t \stackrel{\text{iid}}{\sim} N(0, 1)$ (you may also take $\delta_0 = 0$ although the value of δ_0 will not matter here). Run the regression we ran previously in 4a of X_t on time, $E(X_t) = \delta_0 + \delta_1 t$. Get the p-value for testing $H_0 : \delta_1 = 0$. (Note: you can get p-values from `summary()$coefficients`.)

Do this procedure $M = 1000$ times. Report the proportion of p-values that are smaller than .05. Provide a comment explaining what this means for the p-value reported in 4b.