Disclaimer: You are to work alone on this midterm. You can ask questions of clarification to me, but you can not discuss any part of this exam with anyone else. Giving or receiving unauthorized assistance, or attempting to give or receive such assistance, in connection with exam will constitute a violation of the USC Honor Code.

Disclaimer: Show all of your work to receive full credit! On "R problems," you must attach all code and output to receive credit.

Note: This exam is due at 2.00pm on Monday, October 21. You must turn in your exam to me directly (that is, you can not have someone else turn it in for you). If necessary, slide under my office door or put in my mailbox. Distance students can email their exam to me (but it might be better to mail your exam to me if you have written work).

1. Suppose that $\{e_t\}$ is a zero mean white noise process with $\text{var}(e_t) = \sigma_e^2$. Let B denote the backshift operator. Consider the processes

(i) $(1 + 0.4B)Y_t = e_t$

(ii) $(1 - 1.5B + 0.56B^2)Y_t = (1 + 0.8B)e_t$

(iii) $(1 - 0.9B)(1 - B)Y_t = (1 - 0.5B)(1 + 0.4B)e_t$,

(iv) $(1 - 0.4B - 0.45B^2)Y_t = (1 + B + 0.25B^2)e_t$

(v) $(1 + 1.8B - 0.6B^2 + 0.4B^3)Y_t = e_t$.

(a) Identify each model as an ARIMA(p, d, q) process; that is, specify p, d, and q.
(b) Give the autocorrelation function ρ_k for those processes which are stationary. If you want, you can just use the `ARMAacf` function in R and list the first dozen or so correlations.
(c) Simulate a data set from each process identified above (only for those that are stationary); show the data set in a time series plot and also show the sample ACF for it. In each case, does the sample ACF agree with what we know to be true from the theory? (Discuss each one). I will let you select the sample size n (e.g., larger than 100) and the white noise variance.

2. Suppose $Y_t = X_t + W_t$, where W_t is a zero mean normal white noise process with variance $\text{var}(W_t) = \sigma_W^2$. The $\{X_t\}$ process is a stationary AR(1) defined by $X_t = \phi X_{t-1} + Z_t$, where Z_t is a zero mean normal white noise process with variance $\text{var}(Z_t) = \sigma_Z^2$. As usual, in the AR(1) process, assume that Z_t is independent of $X_{t-1}, X_{t-2}, \ldots,$. Assume additionally that $E(W_t Z_s) = 0$ for all t and s.
(a) Show that $\{Y_t\}$ is stationary and find its autocovariance function.
(b) Show that the process $\{U_t\}$, where

$$U_t = Y_t - \phi Y_{t-1} = (1 - \phi B)Y_t,$$

has nonzero correlation only at lag 1 (excluding lag 0, of course!).
(c) Show that $\{Y_t\}$ behaves like an ARMA(1,1) process.

3. (a) Describe, in detail, the important characteristics of the autocorrelation function for the MA(1), MA(2), AR(1), AR(2), and the ARMA(1,1) processes. Be thorough.
(b) In Figure 1 (next page), I have displayed four sample autocorrelation functions (ACFs). Each ACF corresponds to a different simulated ARMA process. Make educated guesses about

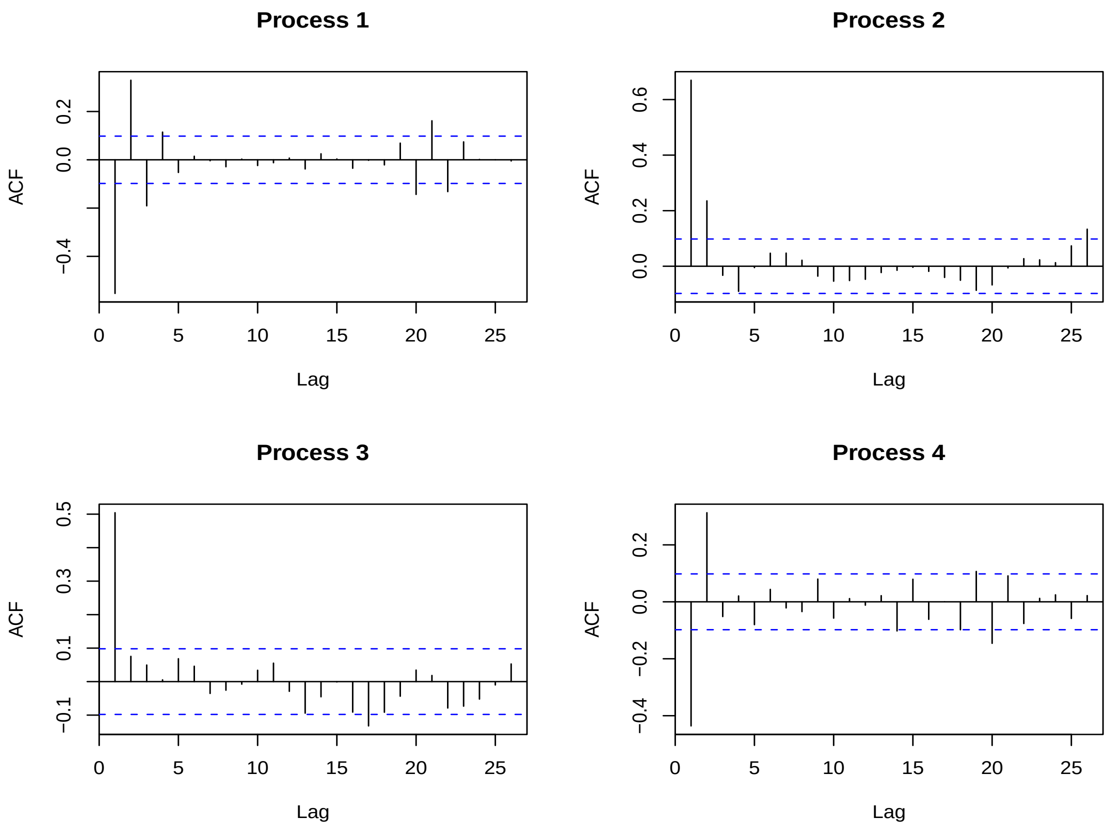

Figure 1: Sample autocorrelation functions for four different simulated processes.

which model was used to produce each sample ACF. Just the model name will be sufficient (don't worry about parameter values), but defend your answers with clear reasoning.
(c) In Figure 1, I placed margin of error cutoffs under the white noise assumption. What sample size n did I use in the four simulations? It was the same in each.

4. The United States is commonly known as a "land of immigrants." For this question, we will focus on legal immigrants who came to the United States and were classified as a "late-life immigrant" (i.e., were aged 65 years or older when entering the United States). The data set `immigrant` (see the course web page) contains the number of late-life immigrants to the United States (in 1000s) during 1933-2007. A student in my Fall, 2011 class collected these data from the Department of Homeland Security and used the data for her class project. I have displayed the raw data in Figure 2.
(a) Use the regression methods in Chapter 3 to detrend the data, that is, fit a deterministic trend model of the form

$$Y_t = \mu_t + X_t,$$

where μ_t is a deterministic trend function. Obviously, you are to examine the data and identify a plausible trend function. Produce a plot that displays the data with your fitted model superimposed over them (like in the notes). **Note:** You might wish to transform the data first;

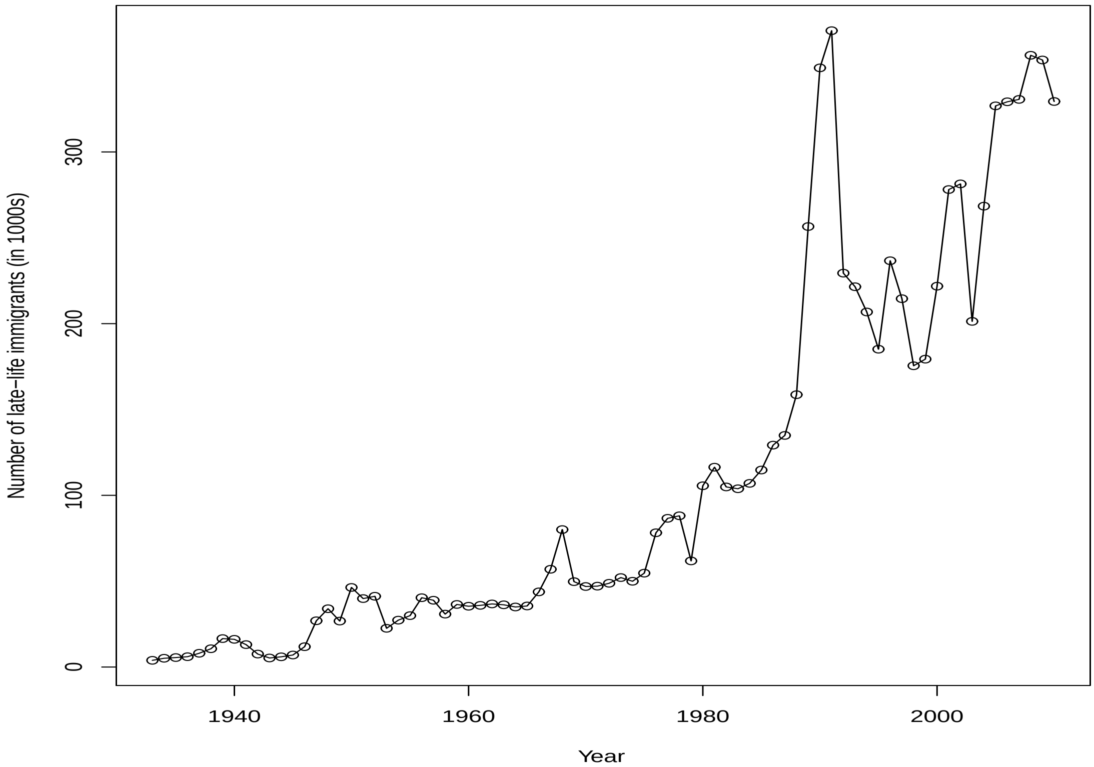

Figure 2: Immigrant data. Number of late-life immigrants (in 1000s) to the United States during 1933-2007.

in this case, fit a deterministic trend model of the form

$$T(Y_t) = \mu_t^* + X_t,$$

where $T(Y_t)$ denotes the transformed data and μ_t^* denotes the deterministic trend function on the transformed scale. In either model, assume that $E(X_t) = 0$.
(b) Examine the standardized residuals $\{\widehat{X}_t^*\}$ from your fitted model (transformed or not) for normality and independence. What are your conclusions? Do the standardized residuals look to resemble a normal, zero mean white noise process?
(c) Display the sample ACF for the standardized residuals in part (b). What type of stationary model would be useful in modeling the standardized residual process?

5. Suppose that $\{Y_t\}$ is an IMA(1,1) process; i.e.,

$$Y_t = Y_{t-1} + e_t - \theta e_{t-1},$$

where $\{e_t\}$ is zero mean white noise with $\text{var}(e_t) = \sigma_e^2$. Recall that this is a nonstationary process.
(a) What is the name of the process identified by $\{\nabla Y_t\}$?
(b) Derive the mean, autocovariance, and autocorrelation functions for $\{\nabla^2 Y_t\}$. Is $\{\nabla^2 Y_t\}$

stationary? Is $\{\nabla^2 Y_t\}$ invertible?
(c) Suppose that you have observed a realization of the $\{Y_t\}$ process, that is, data $Y_1, Y_2, ..., Y_n$ that follow an IMA(1,1) model, and that you would like to estimate θ using the data. Describe an approach to do this. *Hint:* Think about the process identified by $\{\nabla Y_t\}$. You should know the true autocorrelation function for this process (and this function depends on θ).
(d) The data set called `robot` can be found in the authors' TSA library. This time series was obtained from an industrial robot. The robot was put through a sequence of maneuvers and the distance from a desired ending point was recorded (in inches). This was repeated $n = 324$ times to form the series.

(d1) Prepare a quality time series plot for these data. Describe any interesting features you see.

(d2) My colleague has suggested to me that an IMA(1,1) model describes these data very well. Prepare a thorough analysis/argument that supports my colleague's claim. Using your approach in part (b), what would you estimate θ to be based on these data and under the IMA(1,1) model assumption?

(d3) My other colleague agrees that an IMA(1,1) model describes these data very well but is convinced that these data display heteroscedasticity (nonconstant variance). What do you think of this colleague's claim?

Request: Please keep problems separate (e.g., do not log all R code at the end of your solutions). Also, neatness and organization counts! You don't want me to have to "search for things" in your solutions. Pay attention to details and explain/present everything as clearly as possible!! You do not have to "type up" all of your solutions; they can be hand written (some of mine are).