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 5.15pm on Wednesday, October 19. 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. You can also email your exam to me.

1. Suppose that $\{X_t\}$ is a stationary AR(1) process with parameter $\phi \neq 0$, that is,

$$X_t = \phi X_{t-1} + e_t,$$

where $\{e_t\}$ is zero mean white noise with $\text{var}(e_t) = \sigma_e^2$. Define the two processes $\{Y_t\}$ and $\{W_t\}$ by

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 t + X_t \\ W_t &= Y_t - Y_{t-1}. \end{aligned}$$

In the definition of Y_t, β_0 and β_1 are fixed constants (i.e., they are not random and they are not functions of t).
(a) We are given that $\{X_t\}$ is stationary. What does this tell us about the value of ϕ? What does this tell us about the value of $E(X_t)$?
(b) Find the mean and variance of Y_t. Is the $\{Y_t\}$ process stationary?
(c) Find the mean and variance of W_t. Is the $\{W_t\}$ process stationary?
(d) Pick (nonzero) numerical values for σ_e^2, ϕ, β_0, and β_1. **Note:** Keep $-0.25 \leq \beta_1 \leq 0.25$! Use R to generate a realization (of length $n = 125$) of each process: $\{X_t\}$, $\{Y_t\}$, and $\{W_t\}$.

(d1) Display each realization in a time series plot (suitably labeled/titled) and describe the appearance of each plot.

(d2) Display the sample autocorrelation function (ACF) for your simulated realizations of $\{X_t\}$, $\{Y_t\}$, and $\{W_t\}$. Interpret each sample ACF.

2. Suppose that $\{e_t\}$ is zero mean white noise with $\text{var}(e_t) = \sigma_e^2$. Consider the models:

(i) $Y_t = 0.80Y_{t-1} + e_t - 0.20e_{t-1}$

(ii) $Y_t = Y_{t-1} + e_t + 0.25e_{t-1} + 0.50e_{t-2}$

(iii) $Y_t = 0.33Y_{t-1} + 0.66Y_{t-2} - Y_{t-3} + e_t - 1.66e_{t-1} + 2.33e_{t-2}$.

(iv) $Y_t = 0.80Y_{t-1} - 0.15Y_{t-2} + e_t - 0.30e_{t-1}$.

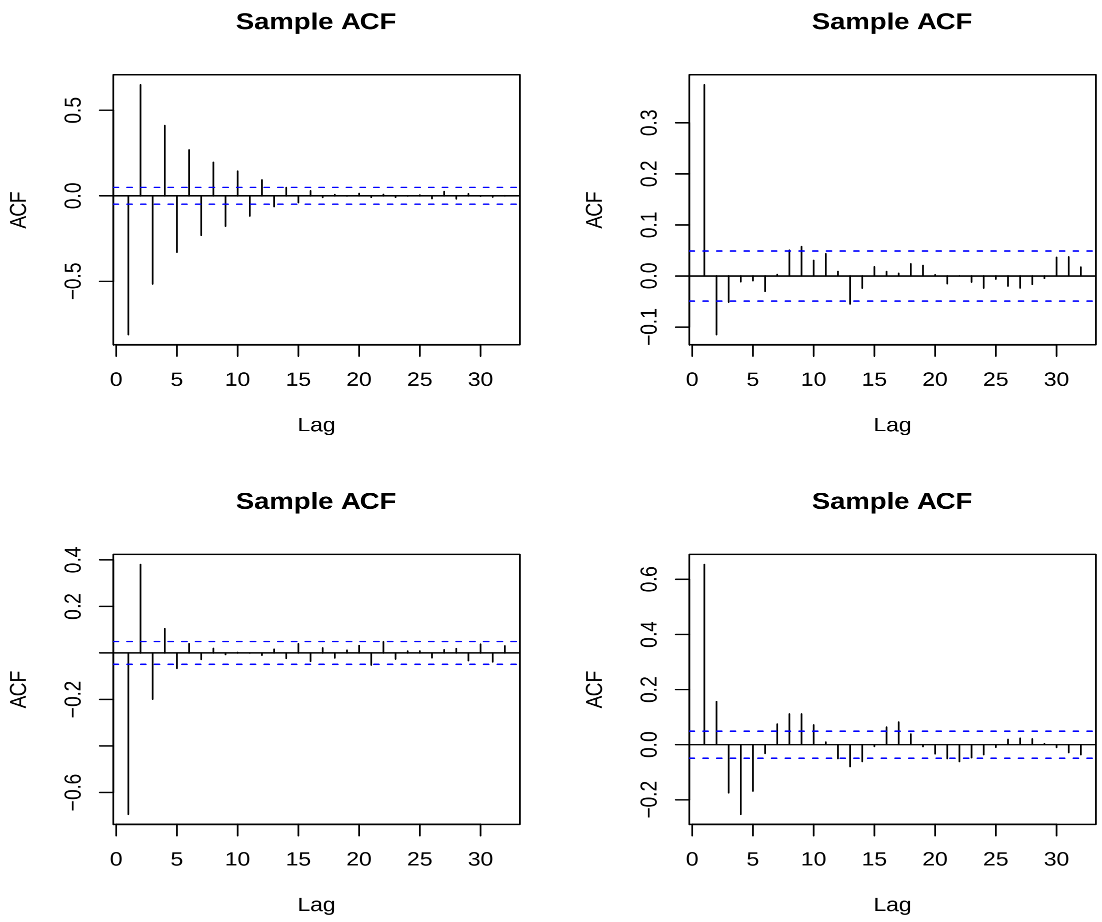

Figure 1: Sample autocorrelation functions for four different simulated processes.

(a) Write each of these models using backshift notation.
(b) Determine whether each model is stationary and/or invertible.
(c) Identify each model as an ARIMA(p, d, q) process; that is, specify p, d, and q.
(d) For the process in (i), find the first 5 (theoretical) autocorrelations ρ_1, ρ_2, ..., ρ_5.

3. (a) Describe the stationarity and invertibility conditions for the MA(1), MA(2), AR(1), AR(2), and the ARMA(1,1) models. Be thorough.
(b) Describe the salient characteristics of the autocorrelation function for the MA(1), MA(2), AR(1), AR(2), and the ARMA(1,1) models. Be thorough.
(c) In Figure 1 (above), I have displayed four sample autocorrelation functions (ACFs). Each ACF corresponds to a different simulated ARMA process. Make educated guesses about which model was used to produce each sample ACF. Just the model name will be sufficient but defend your answers with clear reasoning.
(d) 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.

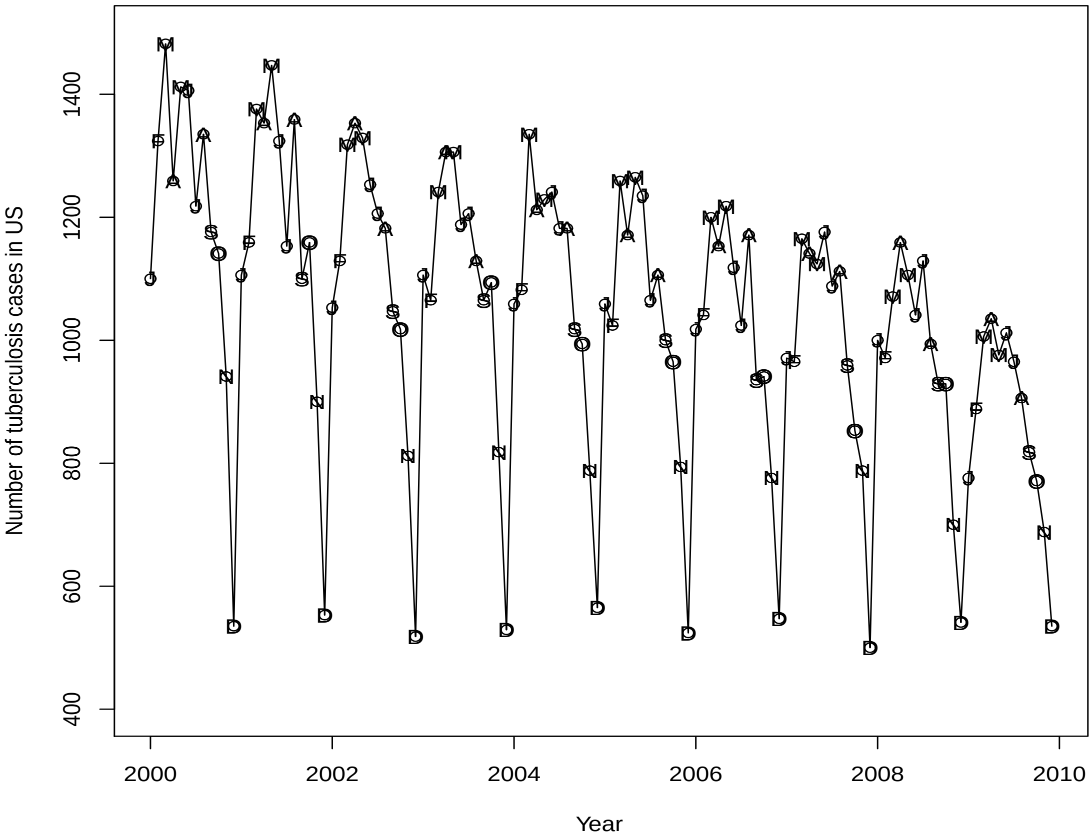

Figure 2: Number of TB cases by treatment start date in the United States.

4. Tuberculosis, commonly known as TB, is a bacterial infection that can spread through the lymph nodes and bloodstream to any organ in your body (it is most often found in the lungs). Most people who are exposed to TB never develop symptoms, since the bacteria can live in an inactive form in the body. But if the immune system weakens, such as in people with HIV or elderly adults, TB bacteria can become active and fatal if untreated. Figure 2 diplays the number of TB cases (per month) in the United States from January 2000 to December 2009. The data can be found on the course home page (under Data sets; filename = `tb`).

(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; 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.
(d) The data `schiz` can be found on the course home page (under Data sets; filename = `schiz`). These data are daily perceptual speed scores for a schizophrenic patient (recorded over $n = 120$ days).

(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" your solutions; they can be hand written (mine are).