Assignment #2

*Submit your* ***printed*** *solution at the start of class on September 26.*

Short answers, please. *Explain your answer concisely*. Include a plot only if requested in the question or if it makes the point better than you can in words. “Significance” implies statistical significance. Presume necessary conditions hold for tests of significance unless conditions are specifically addressed by the question.

You will need to download two data files for this assignment from the Canvas Assignments folder. The first three questions use data in the file a2\_series.csv, and the remaining seven questions refer to the file unrate\_quarterly.csv.

Once you have downloaded the first CSV file, create two time series as follows:

Data <- read.csv("[your file path]a2\_series.csv")  
xt <- ts(Data$xt, start=1)  
yt <- ts(Data$yt, start=1)

1. Detrend the nonstationary time series xt. Justify your choice of method. Your answer should include *relevant* key plots.
2. Repeat Question 1, but with the time series yt.
3. The two time series xt and yt are contemporaneously correlated, but might this correlation be spurious in the sense that the underlying processes *Xt* and *Yt* are in fact independent of one another? Explain.

The remaining questions concern regression models for an updated version of the Hawaiian occupancy data hor discussed in the textbook (*e.g*. Example 3.20). First extend this time series with 4 more years of data. Then define economic time series to use as predictors. All of these time series have length 152, spanning 1982-2019.

library(astsa)  
# occRate Add 4 years of data to hor   
occRate <- c(hor, c(80.7, 77.5, 80.5, 77.5, # 2016  
 81.4, 79.4, 81.4, 78.6, # 2017  
 83.5, 81.1, 80.1, 76.4, # 2018  
 80.5, 80.3, 82.8, 79.6)) # 2019  
occRate <- ts(occRate, start=1982, freq=4)  
  
# unRate FRED data [Canvas assignments folder]  
u <- read.csv("[your file path]unrate\_quarterly.csv")  
u <- ts(u[,"UNRATE"], start=1948, freq=4)  
unRate <- window(u, 1982, 2019.75)  
  
# pctChgGNP use a portion of the built-in GNP series  
x <- window(GNP, start=1981.75, end=2019.75)  
pctChgGNP <- diff(x)/x[-length(x)]  
  
# quarter categorical feature identifying quarter  
quarter <- as.factor(rep(1:4, length(occRate)/4))

Once you have created these time series, fit the following regression model. Show a summary of this model as part of your solutions.

regr\_1 <- dynlm(occRate ~ L(occRate,1) + L(occRate, 4)   
 + quarter   
 + L(unRate,1) + L(unRate,2))

1. Briefly interpret the coefficient of the predictor quarter2 in this regression.

The remaining questions refer to the following regression that adds two predictors built from the data on GNP. Show a summary of this model as part of your solutions.

regr\_2 <- dynlm(occRate ~ L(occRate,1) + L(occRate, 4)   
 + quarter   
 + L(unRate,1) + L(unRate,2)   
 + L(pctChgGNP,1) + L(pctChgGNP,2))

1. Assuming the required conditions for the regression model hold, is regr\_2 significantly more predictive of the occupancy rate than regr\_1? Provide evidence to support your answer.
2. The two predictors based on the percentage change in GNP have p-values greater than 0.05 in the estimated regression. Should this apparent lack of significance be attributed to collinearity? Explain briefly.
3. Diagnostic plots reveal 3 rather large negative residuals. Are these outliers influential, having undue influence of the coefficient of any of these predictors?
4. Do the residuals from regr\_2 by-and-large meet the conditions required for inference on coefficients of the regression?
5. Generate a 95% prediction interval for the occupancy rate in the first quarter of 2020, assuming regr\_2 meets the standard conditions that assume normally distributed, independent errors with constant variance.   
   [You’ll need to compute this interval “by hand”; the dynlm package doesn’t supply a convenient predict function! Maybe one of you should contribute to this open-source community and write one. I think there might be a sneaky way to do this.]
6. The estimated coefficients for the predictors derived from the unemployment rate suggest a simplification of regr\_2 that uses differences of unemployment. Make that change and test whether this restriction significantly degrades the apparent predictive accuracy of the model.