Econ 108 FAll 2022
Problem Set 4

This problem set is due at 7pm, Friday, October 28, 2022.

1. Familiarize with the dja.R code posted on canvas (in the modified_code directory under the files tab) before working on this question.

   (a) Download the nyse5.csv dataset from canvas to the working directory on your laptop. This data set contains weekly stock price index from the New York Stock Exchange between January 1976 and March 1989.

   (b) Plot the weekly price index against time (in weeks). By looking at the plot, does the price index time series appear to be stationary or to be a unit root (random walk) process?

   (c) Plot the autocorrelation function over time for the weekly price index series. Does the acf plot suggest evidence of persistence over time?

   (d) Run an AR(1) regression of price on its own lag (lag 1):

   $$p_t = \alpha + \beta p_{t-1} + \epsilon.$$

   Using the estimated coefficient and std error, can you reject the null hypothesis that the true β is equal to 1 (namely that the true model is a unit root, or random walk process)?

   (e) Compute the weekly return time series from the weekly price index series:

   $$r_t = \frac{p_t - p_{t-1}}{p_{t-1}}.$$

   (f) Plot both the return series and the autocorrelation function over time for the return series. By looking at the plots, does the return time series appear to be stationary or to be a unit root (random walk) process?

   (g) Run an AR(1) regression of returns on its own lag (lag 1):

   $$r_t = \alpha + \beta r_{t-1} + \epsilon.$$

   The *efficient market hypothesis* in financial economics conjectures that returns are not predictable. Can you test this hypothesis based on the output of this regression?

(h) Use the AER package to calculate heteroscedasticity robust std errors for the returns AR(1) regression:

$$r_t = \alpha + \beta r_{t-1} + \epsilon.$$

Do the resulting standard deviations change visibly?

2. Familiarize with the airline.R code posted on canvas (in the modified_code directory under the files tab) before working on this question.

(a) Download the ezanders5.csv dataset from canvas to the working directory on your laptop. The data are on monthly unemployment claims in Anderson Township in Indiana, from January 1980 through November 1988.

(b) Plot the unemployment claims against time, and then the log of unemployment claims against time. Which of these two is better modeled by a linear regression model? Are there visible time trend and seasonality in the data plot?

(c) We will choose to work with the log of the unemployment claims. Run a linear regression of the log of the unemployment claims on a time trend and the month dummies. Overlay the fitted value of the regression on the data.

(d) Plot the residual of this regression and its autocorrelation function over time. Do you see evidence of serial persistence over time?

(e) Augment the previous regression with the lag of log unemployment:

$$\log(unempl)(t) = \alpha + \beta_t t + \beta_m' 1_m + \log(unempl)(t-1) + \epsilon_t.$$

(f) Plot the residual of this regression and its autocorrelation function over time. Does the augmented regression help to reduce serial persistence of the residuals over time?

(g) The last column of the dataframe is variable called "ez" that is 1 after 1984 and 0 before 1984. In 1984, an enterprise zone (EZ) was established in Anderson (and other cities in Indiana). Can you augment the regression model to see whether creating an enterprise zone has an effect on the unemployment claims?