Problem Set #6
Spring 2017

1. Rewrite the following expressions without using the lag operator
   (a) $(1-\rho L)y_t = \varepsilon_t$
   (b) $(1+.2L-.8L^2)y_t = \varepsilon_t$
   (c) $y_t = (1-\theta L)\varepsilon_t$
   (d) $y_t = (1-.3L+.5L^2)\varepsilon_t$
2. Rewrite the following expressions in lag operator form
   (a) $y_t = .8y_{t-1} + \varepsilon_t$
   (b) $y_t = .2y_{t-1} + .3y_{t-2} - .7y_{t-5} + \varepsilon_t$
   (c) $y_t = \varepsilon_t - .7\varepsilon_{t-1}$
   (d) $y_t = \varepsilon_t + .3\varepsilon_{t-4}$
3. Consider the following MA(1) process

$$y_t = 2.3 - 0.95\varepsilon_{t-1} + \varepsilon_t.$$

   with ε_t WN(0,1)
   (a) What is the optimal forecast for time periods $T+h$, $h = 1, 2, 3$. Write your answer as a function of $y_1, ..., y_T$ and/or $\varepsilon_1, ..., \varepsilon_T$
   (b) Now suppose that $\varepsilon_T = 0.4$ and $\varepsilon_{T-1} = -1.2$. Re-answer part (a)
4. The stata file "realgdpgrowth.dta" contains quarterly observations on the real growth rates of the components of U.S. gdp.

   Pick two of the series. Graph their autocorrelation functions. Comment and discuss.
5. This question concerns "housing starts", which is the number of new homes where construction has started, monthly from 1959:1 to 2017:1. The FRED labels for the midwest region (of the U.S.), seasonally adjusted, is HOUSTMW and seasonally unadjusted is HOUSTMWNSA. The labels for the south region (of the U.S.) are HOUSTS and HOUSTSNSA. Graph the autocorrelation functions for the series. Comment and discuss.
6. The file "s&p.dta" has stock price and transaction volume data. Estimate a trend model for the natural log of volume. That is, take the log of the volume and regress it on a time trend. Take the residuals of that regression. (The command is **predict e, residuals** after estimating the regression). Graph a time-series plot of the residuals. Graph the autocorrelation function of the residuals. Comment and discuss.
7. For this problem you will create simulated data. See the description of simulation in STATA at the end of this problem set.

   Simulate $T = 100$ observations from each of the four models given below, with ε_t iid $N(0, 1)$. For each series, graph the autocorrelation function, and estimate a MA(2) model. Comment and discuss. For estimation, use the command (where y is the name of the series you created)

   .arima y arima(0,0,2)
   (a) $y_t = \varepsilon_t$ (white noise)
   (b) $y_t = \varepsilon_t + 0.8\varepsilon_{t-1}$

(c) $y_t = \varepsilon_t - 0.6\varepsilon_{t-1} + 0.4\varepsilon_{t-2}$

(d) $y_t = .5y_{t-1} + \varepsilon_t$

Simulation in STATA

Starting with a blank STATA session, you can set the number of observations and declare the file to be time-series as follows. If you want to create a file with 100 observations, use the commands

```
.set obs 100
.gen t=_n
.tsset t
```

Before generating random numbers you can set the "seed" for the random number generator. The command takes the form

```
.set seed 83346563
```

where in this example the "seed" is the number 83346563. Any integer between 1 and about 2 million is appropriate. If you don't specify a seed the default is 123456789 when Stata is started up. You can set the seed to any number you want, perhaps the current time. You can also use an online random number generated such as one as `numbergenerator.org`. (That's how I obtained the seed given above.)

To simulate a Gaussian white noise process use the command

```
.gen e=rnormal()
```

This generates (creates) a variable "e" (or any name you decide) filled with random normal N(0,1) innovations.

To create a moving average process $y_t = \varepsilon_t + 0.3\varepsilon_{t-1}$, use the command

```
.gen y=e+0.3*L.e
```

This takes the variable "e" previously defined, creates the lagged value, and creates a moving average process with the coefficient 0.3

To create an autoregression, you need a bit more creativity because of the recursion. To create the process $x_t = 0.5x_{t-1} + \varepsilon_t$ with $x_1 = 0$

```
.gen x=0
.replace x=0.5*L.x+e if t>1
```

The first command sets the variable $x_t = 0$. We only use this for the first observation, thus $x_1 = 0$. (You can replace 0 with whatever starting value is appropriate.) Then the second command replaces the x_t by the recursion $x_t = 0.5x_{t-1} + e_t$, for observations $t > 1$, where the e variable was previously defined, and the time index t was previously defined. The command is **replace** rather than **gen** since the variable x already exists, and **gen** will not write over an existing variable.