ECN 480/PUB 580
Assignment #6
Due: Thursday, April 21, 2022 by end of day

Directions: Answer each question in a MS Word or .pdf file. Compile your answers into a single computer file, and then upload it to Canvas under "Assignment #6." Contact me if you have any questions.

1. Download the data set entitled "assign 6.dta" from Canvas that is posted along with assignment #6. This is monthly data on traffic accidents for the State of California for January 1981 – December 1989.

The variables included are:

- ltotacc: the natural logarithm of the total number of traffic accidents in a month.
- feb, mar, apr, …, dec: dummy variables that equal 1 if the month is a particular month, zero otherwise. For example, feb = 1 if the month is February, 0 if not.
- t: a time trend. t =1 for the first month in the data set, =2 for the second month, =3 for the third month, etc. There are 108 months in the data set, so t = 1, 2, … 108.
- unem: the state unemployment rate that month
- spdlaw: a dummy variable that = 1 when the speed limit was raised to 65 mph, 0 when the speed limit was still 55 mph.
- beltlaw: a dummy variable that = 1 when a seatbelt law was signed into law. Click-It or Ticket!
- wkends: the number of weekends for each month in the data set.

IMPORTANT: Before using time series data, you have to tell Stata that you have time series data. This is easy to do. You just type the following command: `tsset timevariable` where "timevariable" is whatever time series variable your data is measured over. In this case, just use t as your time series variable. Thus, the command will be:

```
tsset t
```

If you executed the command correctly, you should then see:

```
        time variable:  t, 1 to 108
                delta:  1 unit
```

The "delta" just means the frequency over which your data is measured. In this case, the "unit" is a month. A more advanced command can tell Stata exactly what unit your data is measured over, sow when you execute the tsset command, the unit will be replaced with the specific time unit, such as month. However, what we did will serve our purposes just fine for this assignment.

1. Suppose you think that the traffic accidents are a function of the month of the, a time trend, the unemployment rate, the seatbelt law, whether or not the speed limit is 65 mph, and the number of weekends in a month. Perhaps you think that the state of the economy impacts the number of accidents, hence the inclusion of the unemployment rate. It also might be the case that the month of the year impacts traffic accidents and the number of weekends in a month impacts accidents, since more weekends means more people on the road. Hence the inclusion of those variables. Finally, from a policy perspective, you might want to know if the increase in the speed limit and the seatbelt law influence traffic accidents. Note that it is hard to believe, but raising the speed limit on interstate highways from 55mph to 65mph was a big deal! I remember that happening when I was a kid in the 1980s. Could you imagine only being able to drive 55mph on I-75?? I guess that's why Sammy Hagar wrote the song "I Can't Drive 55! Thus:

ltrotacc = f(feb, mar, apr, may, jun, jul, aug, sep, oct, nov, dec, t, wkends, unem, spdlaw, beltlaw)

Estimate this regression in Stata. Note that there is a shortcut to estimating this regression. Rather than typing each monthly dummy variable separately in the `reg` command, you can just type them as: `feb - dec`. This tells Stata to include all the variables from February to December in the variable list in the regression. You can use this trick for any variables in variable list if you want to include variables starting with one variable on the list and ending with another on the list. **(3 points)**

2. Look at the unemployment, speed limit, and seatbelt law variables. Which ones have a $\hat{\beta}$ that is statistically different from zero? Use a two-tailed test and a 5% level of significance. **(3 points)**

3. Note that we have a log-linear regression, which means that we are using the natural logarithm of the y-variable but not of any of the x-variables. We called this a "Case 2" regression in Lecture #12. What impact did raising the speed limit have on traffic accidents? **(2 points)**

4. What impact does the seatbelt law have on monthly traffic accidents? Is the direction of the effect on the seatbelt law on traffic accidents what you would have expected? Any idea on what might cause the direction of the effect you actually found in the regression? **(2 points)**

5. We have not discussed a time trend yet, but it is straightforward to interpret. A positive $\hat{\beta}$ on the time trend t means that the y-variable is trending upward over time. A negative $\hat{\beta}$ means that the y-variable is trending down over time. The magnitude of the $\hat{\beta}$ just says how much the y-variable is trending up or down each month (or whatever time period your data is measured over). That is, if $\hat{\beta}$ is 0.025, that would mean that traffic accidents are increasing by 2.5% each month. You would include a time trend in a time series model when you think your y-variable is trending up or down over time. Is this the case for traffic accidents in your model? Is the $\hat{\beta}$ on the trend statistically significant? How much are accidents increasing or decreasing each month? **(3 points)**

6. Like any time series model, serial correlation is a concern. Test for serial correlation by generating a variable that is equal to the residuals from your regression, using the `predict uhat, resid` command. Generate a variable that is equal to the lag of the residual from the regression using the command `gen laguhat = uhat[_n-1]` command like we discussed in Lecture #21. Then, estimate the following regression:

$$\hat{u}_t = \hat{\rho}\hat{u}_{t-1} + e_t$$

Test for the statistical significance of $\hat{\rho}$. Is there statistically significant serial correlation in your model? Refer to the end of Lecture #21 for an example of this test. **(3 points)**

7. Estimate the serial correlation and heteroscedasticity-robust standard errors by using the `newey` command in place of the `reg` command for your model in question #1. Use four lags in the model. Refer to Lecture #24 for an example of using this command. Is there any change to the statistical significance that you found in question #5? **(3 points)**

8. Generate a variable that is a one period lag of the ltotacc variable, using the `gen lagltotacc = ltotacc[_n-1]` like we talked about in Lecture #23. . Re-estimate your model from question #1 (using the `reg` command) that uses this new variable as an x-variable along with all the other x-variables. Repeat the test for serial correlation from question #6. Note that you have to drop `uhat` and `laguhat` and then regenerate them after you run this regression. Otherwise, they are the residuals and lag of the residuals for the original regression, not this new regression. Do you still have statistically significant serial correlation? **(3 points)**

9. Refer to your results in question #8. Does the seatbelt law and the speed limit explain traffic accidents over and above just knowing what last month's traffic accidents were? If so, how does the speed limit increase and the seatbelt law impact traffic accidents? **(3 points)**

Clear the data from Stata, download the data set entitled "assign 6_2.dta" from Canvas that is posted along with assignment #6 and load that data set into Blackboard. This is time series data for the yearly inflation rate and yearly unemployment rate between 1948 and 2017. Recall that the "Phillips Curve" says there is a negative relationship between the yearly inflation rate and the yearly unemployment rate. If you take Intermediate Macroeconomics (ECN 354) course semester, you will hear all about it! For our purposes though, it is enough to know that the Phillips Curve says the higher the unemployment rate, the lower the inflation rate and the lower the unemployment rate, the higher the inflation rate. That is:

inf = f(unem)

Where "inf" is the yearly inflation rate and "unem" is the yearly unemployment rate.

10. Estimate a simple regression using the `reg` command for the inflation rate is the y-variable and the unemployment rate is the x-variable. Is there a statistically significant negative relationship between the two as the Phillips Curve suggests? The sign and significance of $\hat{\beta}_1$ will tell you that. **(3 points)**

11. Conduct a Durbin-Watson test for serial correlation like we did in Lecture #22 by typing the command `estat dwatson` after estiming the model in question #10. To save you having to look-up the critical values, d_L = 1.58 and d_U = 1.64. Does the test suggest there is serial correlation, no serial correlation, or is the test inconclusive? **(3 points)**

12. Estimate the Cochrane-Orcutt method to correct for serial correlation like we did in Lecture #23 by executing the command `prais inf unem, corc`. Does this seem to have fixed the serial correlation? How does the transformed Durbin-Watson compare to the critical values? **(2 points)**

13. Refer to your results in question #12. Is there a statistically significant negative relationship between the unemployment rate and the inflation rate after this transformation? **(2 points)**