ECN 480 Assign 6, Winter 2022

Saturday, April 16, 2022 11:17 AM

Casson Konzer :)

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

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)**

```
. reg ltotacc feb - dec t wkends unem spdlaw beltlaw

      Source |       SS           df       MS      Number of obs   =       108
-------------+----------------------------------   F(16, 91)       =     57.61
       Model |  1.14490901        16  .071556813   Prob > F        =    0.0000
    Residual |  .113028469        91  .001242071   R-squared       =    0.9101
-------------+----------------------------------   Adj R-squared   =    0.8943
       Total |  1.25793748       107  .011756425   Root MSE        =    .03524

------------------------------------------------------------------------------
     ltotacc | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         feb |  -.0338346   .0177683    -1.90   0.060    -.0691292    .0014599
         mar |    .076953   .0167941     4.58   0.000     .0435937    .1103124
         apr |   .0104561   .0170469     0.61   0.541    -.0234054    .0443177
         may |   .0237074   .0169389     1.40   0.165    -.0099397    .0573544
         jun |   .0219334   .0172149     1.27   0.206     -.012262    .0561288
         jul |   .0499293   .0167036     2.99   0.004     .0167496    .0831089
         aug |   .0559526   .0168173     3.33   0.001      .022547    .0893581
         sep |   .0420693   .0172819     2.43   0.017      .007741    .0763977
         oct |   .0817171   .0169554     4.82   0.000     .0480372    .1153969
         nov |   .0768721   .0172455     4.46   0.000     .0426161    .1111282
         dec |   .0990863   .0170705     5.80   0.000     .0651779    .1329948
           t |   .0011011   .0002579     4.27   0.000     .0005889    .0016133
      wkends |   .0033333   .0037761     0.88   0.380    -.0041675    .0108342
        unem |  -.0212173   .0033974    -6.25   0.000    -.0279659   -.0144688
      spdlaw |  -.0537593   .0126036    -4.27   0.000    -.0787948   -.0287238
     beltlaw |   .0954528   .0142351     6.71   0.000     .0671766     .123729
       _cons |   10.63986    .063086   168.66   0.000     10.51455    10.76518
------------------------------------------------------------------------------
```

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)**

Unemployment, Speed Limit, & Seat Belt Law all have coefficients significantly different from 0 as their p-values are all below 0.05.

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)**

Raising the speed limit corresponds to a 5.37 percent decrease in the amount of traffic accidents per month.

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)**

Implementing the seatbelt law corresponds to

Implementing the seatbelt law corresponds to a 9.54 percent increase in the amount of traffic accidents per month.

yes I would expect an increase in accidents after the law is implemented.

I suspect drivers are less cautious when wearing a seatbelt due to an increased sense of saftey & a decreased risk in the case of an accident.

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)**

yes the time trend's coefficient is statistically significant.

Accidents are expected to increase by 0.11 percent month over month.

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)**

```
. predict uhat, resid

. gen laguhat = uhat[_n-1]
(1 missing value generated)

. reg uhat laguhat, noc

      Source |       SS           df       MS      Number of obs   =       107
-------------+----------------------------------   F(1, 106)       =     15.46
       Model |  .013370296         1  .013370296   Prob > F        =    0.0002
    Residual |  .091680915       106  .000864914   R-squared       =    0.1273
-------------+----------------------------------   Adj R-squared   =    0.1190
       Total |  .105051212       107  .000981787   Root MSE        =    .02941
```

uhat	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
laguhat	.3553388	.0903771	3.93	0.000	.1761573	.5345203

The coefficient on laguhat ($\hat{u}_{t-1}$) is statistically significant.

yes there is significant serial correlation as the coefficient ($\hat{\rho}$) is 0.355 which is NOT $\emptyset$!

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)**

```
. newey ltotacc feb - dec t wkends unem spdlaw beltlaw, lag(4)

Regression with Newey-West standard errors      Number of obs     =        108
Maximum lag = 4                                 F( 16,         91) =      37.42
                                                Prob > F          =     0.0000
```

ltotacc	Coefficient	Newey-West std. err.	t	P>|t|	[95% conf.	interval]
feb	-.0338346	.0168283	-2.01	0.047	-.067262	-.0004073
mar	.076953	.0183367	4.20	0.000	.0405293	.1133767
apr	.0104561	.0176927	0.59	0.556	-.0246883	.0456006
may	.0237074	.018769	1.26	0.210	-.013575	.0609897
jun	.0219334	.0200354	1.09	0.277	-.0178645	.0617313
jul	.0499293	.0200711	2.49	0.015	.0100605	.089798
aug	.0559526	.0194475	2.88	0.005	.0173225	.0945826
sep	.0420693	.0187812	2.24	0.028	.0047628	.0793759
oct	.0817171	.021099	3.87	0.000	.0398065	.1236277
nov	.0768721	.0210425	3.65	0.000	.0350738	.1186705
dec	.0990863	.0241168	4.11	0.000	.0511813	.1469913
t	.0011011	.0003897	2.83	0.006	.000327	.0018752
wkends	.0033333	.0028746	1.16	0.249	-.0023768	.0090434
unem	-.0212173	.0051302	-4.14	0.000	-.0314079	-.0110268
spdlaw	-.0537593	.0170449	-3.15	0.002	-.0876169	-.0199017
beltlaw	.0954528	.0188985	5.05	0.000	.0579132	.1329924
_cons	10.63986	.0635822	167.34	0.000	10.51356	10.76616

There is no change in the conclusion from Q#5, athough we do see that the p-value on t has increased & thus could no longer reject H_0 ($\beta_t = 0$) at the previously acceptable 99.9% confidence.

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)**

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)**

```
. reg ltotacc lagltotacc feb - dec t wkends unem spdlaw beltlaw

      Source |       SS           df       MS      Number of obs   =       107
-------------+----------------------------------   F(17, 89)       =     67.82
       Model |  1.16545395        17  .068556115   Prob > F        =    0.0000
    Residual |  .089969838        89  .001010897   R-squared       =    0.9283
-------------+----------------------------------   Adj R-squared   =    0.9146
       Total |  1.25542379       106  .011843621   Root MSE        =    .03179

------------------------------------------------------------------------------
     ltotacc | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  lagltotacc |   .3326589   .0914763     3.64   0.000     .1508974    .5144205
         feb |   .0146528   .0193721     0.76   0.451    -.0238392    .0531448
         mar |   .1430088   .0215594     6.63   0.000     .1001708    .1858468
         apr |    .038098   .0164358     2.32   0.023     .0054405    .0707555
         may |   .0722441   .0186011     3.88   0.000     .0352841     .109204
         jun |   .0627452   .0178428     3.52   0.001      .027292    .0981984
         jul |   .0893951   .0173175     5.16   0.000     .0549856    .1238045
         aug |   .0938237    .017187     5.46   0.000     .0596736    .1279739
         sep |   .0761947   .0171704     4.44   0.000     .0420774     .110312
         oct |    .118897   .0172177     6.91   0.000     .0846858    .1531081
         nov |   .0991804   .0162776     6.09   0.000     .0668372    .1315235
         dec |   .1255029   .0163681     7.67   0.000     .0929797     .158026
           t |   .0008893   .0002615     3.40   0.001     .0003696    .0014089
      wkends |   .0032099   .0034091     0.94   0.349     -.003564    .0099838
        unem |  -.0125834   .0036586    -3.44   0.001    -.0198528   -.0053139
      spdlaw |  -.0414191   .0121314    -3.41   0.001    -.0655238   -.0173143
     beltlaw |   .0652123   .0151074     4.32   0.000     .0351943    .0952304
       _cons |   7.018409   .9900325     7.09   0.000     5.051236    8.985583
------------------------------------------------------------------------------

. predict luhat, resid
(1 missing value generated)

. gen llaguhat = luhat[_n-1]
(2 missing values generated)

. reg luhat llaguhat, noc

      Source |       SS           df       MS      Number of obs   =       106
-------------+----------------------------------   F(1, 105)       =      0.14
       Model |  .000122525         1  .000122525   Prob > F        =    0.7057
    Residual |  .089752007       105  .000854781   R-squared       =    0.0014
-------------+----------------------------------   Adj R-squared   =   -0.0081
       Total |  .089874532       106  .000847873   Root MSE        =    .02924

------------------------------------------------------------------------------
       luhat | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    llaguhat |   -.038073   .1005614    -0.38   0.706    -.2374676    .1613217
------------------------------------------------------------------------------
```

The coefficient on llaguhat is not statistically significant & thus we no longer have statistically significant serial correlation.

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)**

The percentage change corresponding to the

the speed limit increase and the seatbelt law impact traffic accidents? **(3 points)**

The percentage change corresponding to the belt law is the largest seen, which is followed by that of the speed limit, and last that of the previous month's accidents at 0.09%. This said we can see that the absolute values of their corresponding t statistics show that we are more confident in the explanation given by the belt law, than that of the speed limit, of which we are more confident in than that of last month's traffic accidents.

— — — — — — — — —

The speed limit increase corresponds to a 4.14 percent decrease in traffic accidents while the implementation of the belt law corresponds to a 6.52 percent increase in traffic accidents.

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)**

```
. tsset year

Time variable: year, 1948 to 2017
        Delta: 1 unit

. reg inf unem

      Source |       SS           df       MS      Number of obs   =        70
-------------+----------------------------------   F(1, 68)        =      0.97
       Model |  7.92839994         1  7.92839994   Prob > F        =    0.3286
    Residual |  556.869471        68  8.18925693   R-squared       =    0.0140
```

```
-------------+----------------------------------   Adj R-squared   =   -0.0005
       Total |  564.797871        69  8.18547639   Root MSE        =    2.8617

------------------------------------------------------------------------------
         inf | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        unem |   .2113274   .2147758     0.98   0.329     -.217251    .6399058
       _cons |   2.298145   1.289437     1.78   0.079    -.2748865    4.871176
------------------------------------------------------------------------------
```

Our Model predicts positive correlation rather than negative, but this prediction is not of statistical significance.

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)**

```
. estat dwatson

Durbin-Watson d-statistic(  2,    70) =  .6785124
```

As our test statistic is less than d_L, we conclude there is serial correlation.

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)**

```
. estat dwatson

Durbin-Watson d-statistic(  2,    70) =  .6785124

. prais inf unem, corc

Iteration 0:  rho = 0.0000
Iteration 1:  rho = 0.6386
Iteration 2:  rho = 0.7477
Iteration 3:  rho = 0.7721
Iteration 4:  rho = 0.7766
Iteration 5:  rho = 0.7774
Iteration 6:  rho = 0.7775
Iteration 7:  rho = 0.7776
Iteration 8:  rho = 0.7776
Iteration 9:  rho = 0.7776

Cochrane-Orcutt AR(1) regression with iterated estimates

      Source |       SS           df       MS      Number of obs   =        69
-------------+----------------------------------   F(1, 67)        =      7.43
       Model |  29.2731936         1  29.2731936   Prob > F        =    0.0082
    Residual |  263.906822        67  3.93890779   R-squared       =    0.0998
-------------+----------------------------------   Adj R-squared   =    0.0864
       Total |  293.180016        68  4.31147082   Root MSE        =    1.9847

------------------------------------------------------------------------------
         inf | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        unem |  -.6448363   .2365388    -2.73   0.008     -1.11697   -.1727029
       _cons |   6.921968   1.751308     3.95   0.000     3.426343    10.41759
-------------+----------------------------------------------------------------
         rho |   .7775634
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    = 0.678512
Durbin-Watson statistic (transformed) = 1.642407
```

```
--------------------------------------------------------------------------------
Durbin-Watson statistic (original)    = 0.678512
Durbin-Watson statistic (transformed) = 1.642407
```

The Cochrane-Orcutt method does seem to have fixed our serial correlation, our test statistic is just barely in the acceptable range, 1.642 > 1.64, but it is not far from being inconclusive.

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)**

There is now a statistically significant relationship between the unemployment & inflation rates. It follows the assumption of the Phillips Curve as it is negative

GREAT Class, Thanks Much :)