Quantitative Methods for Economics

Tutorial 9

Katherine Eyal

TUTORIAL 9

4 October 2010

ECO3021S

Part A: Problems

1. In Problem 2 of Tutorial 7, we estimated the equation

$$\begin{aligned} \widehat{sleep} &= \underset{(112.28)}{3,638.25} - \underset{(0.017)}{0.148}\ totwrk - \underset{(5.88)}{11.13}\ educ + \underset{(1.45)}{2.20}\ age \\ n &= 706,\ R^2 = 0.113 \end{aligned}$$

   where we now report standard errors in parentheses along with the estimates.

   (a) Construct a 95% confidence interval for β_{totwrk}.

   (b) Can you reject the hypothesis $H_0 : \beta_{totwrk} = -0.2$ against the two sided alternative at the 5% level?

   (c) Can you reject the hypothesis $H_0 : \beta_{totwrk} = -1$ against the two sided alternative at the 5% level?

2. Consider the following demand function for chicken:

$$\log Y_t = \beta_0 + \beta_1 \log X_{1t} + \beta_2 \log X_{2t} + \beta_3 \log X_{3t} + \beta_4 \log X_{4t} + u_t$$

   where Y = per capita consumption of chicken, kg

   X_1 = real disposable per capita income, R

   X_2 = real retail price of chicken per kg, R

   X_3 = real retail price of pork per kg, R

   X_4 = real retail price of beef per kg, R.

   The following regression results are obtained using annual data for 1960 – 1982 (standard errors in parentheses):

$$\widehat{\log Y_t} = \underset{(0.1557)}{2.1898} + \underset{(0.0833)}{0.3425}\ \log X_{1t} - \underset{(0.1109)}{0.5046}\ \log X_{2t} + \underset{(0.0997)}{0.1485}\ \log X_{3t} + \underset{(0.1007)}{0.0911}\ \log X_{4t}$$

$$R^2 = 0.9823$$

$$\widehat{\log Y_t} = \underset{(0.1162)}{2.0328} + \underset{(0.0247)}{0.4515}\ \log X_{1t} - \underset{(0.0635)}{0.3772}\ \log X_{2t}$$

$$R^2 = 0.9801$$

(a) In the first regression, is the estimated income elasticity equal to 1? Is the price elasticity equal to -1? Show your work and use a 5% significance level.

(b) Are chicken and beef unrelated products in the sense that chicken consumption is not affected by beef prices? Use the alternative hypothesis that they are competing products (substitutes). Show your work and use a 5% significance level. (*Note*: this is a one-sided alternative.)

(c) Are chicken and beef and pork unrelated products in the sense that chicken consumption is not affected by the prices of beef and pork? Show your work and use a 5% significance level. (*Note*: this is a test of the joint significance of X_3.and X_4.)

(d) Should we include the prices of beef and pork in the demand function for chicken?

(e) Using the second regression, test the hypothesis that the income elasticity is equal in value but opposite in sign to the price elasticity of demand. Show your work and use a 5% significance level. (*Note*: $\text{cov}\left(\widehat{\beta}_1, \widehat{\beta}_2\right) = -0.00142$.)

(f) Suppose that the demand equation contains heteroskedasticity. What does this mean about the tests computed above?

Part B: Computer Exercises

1. The Capital Asset Pricing Model (CAPM), developed by John Lintner and William F. Sharpe in the 1960s, gives a convenient tool for assessing the performance of asset prices. According to the CAPM, when markets are in equilibrium, the riskiness of an asset relative to the riskiness of the entire asset market would be equal to the slope, β, in the relationship

$$\begin{aligned} & \text{(asset's excess return above the riskless rate)} \\ = \; & \beta \, \text{(excess return of a “market portfolio” above the riskless rate)} + u \end{aligned}$$

where a “market portfolio” is a portfolio containing every asset in the marketplace in proportion to its total value, and u is a mean zero, serially uncorrelated, homoskedastic disturbance term. The coefficient β, usually called the asset's "beta", measures the marginal contribution of the asset to a market portfolio's undiversifiable risk. If $\beta = 0.5$, then when the market excess return rises by 10%, this asset's excess return would rise by 5%.

In 1972, Black, Jensen and Scholes proposed that the validity of the CAPM can be tested by asking whether $\beta_0 = 0$ in

$$\text{(asset's excess return)} = \beta_0 + \beta_1 \, \text{(market's excess return)} + v$$

The data set CAPM2.DTA contains monthly observations for 16 years on the excess returns for six shares (two from each of three industries: the computer, paper, and airline industries). The excess returns for the market are given in *mreturn* and excess returns for the six firms are given in *freturn*.[1] The first 192 observations pertain to firm 1, the next 192 pertain to firm 2, etc. The variable *firm* identifies firms 1 to 6. For each of these 6 firms, test the null hypothesis that $\beta_0 = 0$. What do you conclude about the CAPM from these data? (You can use the `test` command to conduct hypothesis tests in Stata. Use the command: `help test` for more information.)

2. Nitrogen dioxide (NO_2) is a pollutant that attacks the human respiratory system; it increases the likelihood of respiratory illness. One common source of nitrogen dioxide is automobile exhaust. The file NO2POLLUTION.DTA contains a subset of 500 hourly observations made from October 2001 to August 2003. The variables in the data set are

lno2	Natural log of the concentration of NO_2 (particles)
lcars	Natural log of the number of cars per hour
temp	Temperature 2 metres above the ground (degrees C)
wndspd	Wind speed (metres/second)
tchng23	The temperature difference between 25 metres and 2 metres above ground (degrees C)
wnddir	Wind direction (degrees between 0 and 360)
hour	Hour of day
days	Day number from October 1, 2001

   (a) Regress NO_2 concentration on the log of the number of cars, the two temperature variables, the two wind variables, and the time index (*days*). Which variables are significant at the 1% level? At the 5% level? At the 10% level? Interpret your results in full.

   (b) Build a 95% confidence interval for the elasticity of NO_2 pollution with respect to car traffic and check that it matches the Stata output. Is NO_2 pollution elastic or inelastic with respect to car traffic?

   (c) Test the hypothesis that, after controlling for *lcars, temp, tchng23* and *days*, the wind variables have no effect on NO_2 pollution.

   (d) Does a temperature increase of 1 degree C have the same effect as a wind speed increase of 1 metre/second on NO_2 pollution?

   (e) What is the estimated rate of change in NO_2 pollution per day?

   (f) Is it correct to estimate the annual growth rate in NO_2 pollution by multiplying your estimate in (e) by 365? Briefly explain your answer.

   (g) How much of the variation in the log of hourly levels of NO_2 pollution in this sample is accounted for by the variation in the regressors?

[1] The excess return is calculated as the share's rate of return less the rate of return on a risk-free asset.

(h) How much of the variation in the log of hourly levels of NO_2 pollution in this sample could be accounted for by the variation in *days* alone?

3. Consider a model where the return to education depends upon the amount of work experience (and vice versa):

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 educ \cdot exper + u.$$

where *wage* denotes monthly earnings, *educ* denotes years of education and *exper* denotes years of work experience.

(a) Show that the return to another year of education (in decimal form), holding *exper* fixed, is $\beta_1 + \beta_3 exper$.

(b) State the null hypothesis that the return to education does not depend on the level of *exper*. What do you think is the appropriate alternative?

(c) Use the data in WAGE2.DTA to test the null hypothesis in (b) against your stated alternative. (In order to estimate the regression model, you will first need to create a new variable: `gen educXexper = educ*exper` and then incorporate this interaction term into the regression: `reg lwage educ exper educXexper`)

(d) Let θ_1 denote the return to education (in decimal form), when $exper = 10$: $\theta_1 = \beta_1 + 10\beta_3$. Obtain $\widehat{\theta}_1$ and a 95% confidence interval for $\widehat{\theta}_1$. (*Hint*: Write $\beta_1 = \theta_1 - 10\beta_3$ and plug this into the equation; then rearrange. This gives the regression for obtaining the confidence interval for θ_1.)

TUTORIAL 9 SOLUTIONS

4 October 2010

ECO3021S

Part A: Problems

1. In Problem 1 of Tutorial 8, we estimated the equation

$$\begin{aligned} \widehat{sleep} &= \underset{(112.28)}{3,638.25} - \underset{(0.017)}{0.148}\ totwrk - \underset{(5.88)}{11.13}\ educ + \underset{(1.45)}{2.20}\ age \\ n &= 706,\ R^2 = 0.113 \end{aligned}$$

where we now report standard errors in parentheses along with the estimates.

(a) Construct a 95% confidence interval for β_{totwrk}.

(b) Can you reject the hypothesis $H_0 : \beta_{totwrk} = -0.2$ against the two sided alternative at the 5% level?

(c) Can you reject the hypothesis $H_0 : \beta_{totwrk} = -1$ against the two sided alternative at the 5% level?

SOLUTION:

(a) Degrees of freedom $= n - k - 1 = 706 - 3 - 1 = 702$. The 97.5^{th} percentile in a t_{702} distribution: $c = 1.96$.
Thus the confidence interval for β_{totwrk} is $-0.148 \pm 1.96\,(0.017)$, or $[-0.181\,32, -0.114\,68]$.

(b) We can reject the null hypothesis that $\beta_{totwrk} = -0.2$.

(c) We can also reject the null hypothesis that $\beta_{totwrk} = -1$.

2. Consider the following demand function for chicken:

$$\log Y_t = \beta_0 + \beta_1 \log X_{1t} + \beta_2 \log X_{2t} + \beta_3 \log X_{3t} + \beta_4 \log X_{4t} + u_t$$

where Y = per capita consumption of chicken, kg

X_1 = real disposable per capita income, R

X_2 = real retail price of chicken per kg, R

X_3 = real retail price of pork per kg, R

X_4 = real retail price of beef per kg, R.

The following regression results are obtained using annual data for 1960 – 1982 (standard errors in parentheses):

$$\widehat{\log Y_t} = \underset{(0.1557)}{2.1898} + \underset{(0.0833)}{0.3425} \log X_{1t} - \underset{(0.1109)}{0.5046} \log X_{2t} + \underset{(0.0997)}{0.1485} \log X_{3t} + \underset{(0.1007)}{0.0911} \log X_{4t}$$

$$R^2 = 0.9823$$

$$\widehat{\log Y_t} = \underset{(0.1162)}{2.0328} + \underset{(0.0247)}{0.4515} \log X_{1t} - \underset{(0.0635)}{0.3772} \log X_{2t}$$

$$R^2 = 0.9801$$

(a) In the first regression, is the estimated income elasticity equal to 1? Is the price elasticity equal to -1? Show your work and use a 5% significance level.

(b) Are chicken and beef unrelated products in the sense that chicken consumption is not affected by beef prices? Use the alternative hypothesis that they are competing products (substitutes). Show your work and use a 5% significance level. (*Note*: this is a one-sided alternative.)

(c) Are chicken and beef and pork unrelated products in the sense that chicken consumption is not affected by the prices of beef and pork? Show your work and use a 5% significance level. (*Note*: this is a test of the joint significance of X_3.and X_4.)

(d) Should we include the prices of beef and pork in the demand function for chicken?

(e) Using the second regression, test the hypothesis that the income elasticity is equal in value but opposite in sign to the price elasticity of demand. Show your work and use a 5% significance level. (*Note*: $\text{cov}\left(\widehat{\beta}_1, \widehat{\beta}_2\right) = -0.00142$.)

(f) Suppose that the demand equation contains heteroskedasticity. What does this mean about the tests computed above?

SOLUTION:

(a) Is the estimated income elasticity equal to 1?

$H_0 : \beta_1 = 1$

$H_1 : \beta_1 \neq 1$

This is a two sided test.

$$\begin{aligned} t\text{-stat} &= (0.3425 - 1)/0.0833 \\ &= -7.89 \end{aligned}$$

Critical value at 5% level

$$\begin{aligned} c &= t_{a,n-k-1} = t_{0.05,23-4-1} \\ &= 2.101 \end{aligned}$$

Since $|t| > c$, we reject the null at the 5% level of signficance and conclude that income elasticity of demand is not equal to unity.

Is the price elasticity equal to -1?

$H_0 : \beta_2 = -1$

$H_1 : \beta_2 \neq 1$

This is a two sided test.

$$\begin{aligned} t\text{-stat} &= (-0.5046 + 1)/0.1109 \\ &= 4.4671 \end{aligned}$$

Critical value at 5% level

$$\begin{aligned} t_{a,n-k-1} &= t_{0.05,23-4-1} \\ &= 2.101 \end{aligned}$$

Since $|t| > c$, we reject the null at the 5% level of signficance and conclude that income elasticity of demand is not equal to unity.

(b) $H_0 : \beta_4 = 0$

$H_1 : \beta_4 > 0$

The alternate is that β_4 is positive because if the beef and chicken are subsititutes, then when price of beef goes up the demand for chicken increases. This is a one sided test.

$$\begin{aligned} t\text{-stat} &= (0.0911)/0.1007 \\ &= 0.904 \end{aligned}$$

Critical value at 5% level

$$\begin{aligned} t_{a,n-k-1} &= t_{0.05,23-4-1} \\ &= 1.734 \end{aligned}$$

Since $t < c$, we cannot reject the null at the 5% level of signficance and conclude that chicken and beef are unrelated products.

(c) Here $H_0 : \beta_3 = \beta_4 = 0$. We are testing exclusion restrictions, and the first regression given above is the unconstrained regression, while the second is the constrained regression. Note that the R^2 values of the two regressions are comparable since the dependent variable in the two models is the same.

Now the R^2 form of the F statistic is

$$\begin{aligned} F &= \frac{\left(R_{ur}^2 - R_r^2\right)/q}{\left(1 - R_{ur}^2\right)/\left(n - k - 1\right)} \\ &= \frac{\left(0.9823 - 0.9801\right)/2}{\left(1 - 0.9823\right)/\left(23 - 4 - 1\right)} \\ &= 1.1224 \end{aligned}$$

which has the F distribution with 2 and 18 df.

At 5%, clearly this F value is not statistically significant $[F_{0.5}(2, 18) = 3.55]$. The p value is 0.3472. Therefore, there is no reason to reject the null hypothesis – the demand for chicken does not depend on pork and beef prices. In short, we can accept the constrained regression as representing the demand function for chicken.

(d) The test from (c) above shows that the demand for chicken does not depend on pork and beef prices. This suggests that the more parsimonius regression is preferred. However if we have a strong theoretical prior that the prices of pork and beef are significant determinants of the demand for chicken, then we might want to control for them in our model to avoid possible bias in the estimated coefficients.

(e) $H_0 : \beta_1 = -\beta_2 \Rightarrow \beta_1 + \beta_2 = 0$

$H_1 : \beta_1 \neq \beta_2 \Rightarrow \beta_1 + \beta_2 \neq 0$

This is a two sided test.

$$t = \frac{\widehat{\beta}_1 + \widehat{\beta}_2}{\text{se}\left(\widehat{\beta}_1 + \widehat{\beta}_2\right)}$$

where

$$\text{se}\left(\widehat{\beta}_1 + \widehat{\beta}_2\right) = \left\{\left[\text{se}\left(\widehat{\beta}_1\right)\right]^2 + \left[\text{se}\left(\widehat{\beta}_2\right)\right]^2 + 2\text{cov}\left(\widehat{\beta}_1, \widehat{\beta}_2\right)\right\}^{1/2}$$

Therefore

$$\begin{aligned} t &= \frac{\left(0.4515 - 0.3772\right)}{\left[\left(0.02472\right)^2 + \left(0.0635\right)^2 + 2\left(-0.00142\right)\right]^{1/2}} \\ &= \frac{0.0743}{0.042466} \\ &= 1.7496 \end{aligned}$$

Critical value at 5% level

$$\begin{aligned} c &= t_{a,n-k-1} = t_{0.05,23-4-1} \\ &= 2.101 \end{aligned}$$

Since $|t| < c$, we cannot reject the null at the 5% level of signficance and conclude that income elasticity is equal in value but opposite in sign to the price elasticity of demand.

(f) We have a biased estimate for the variance and covariance of our slope estimators, thus our hypothesis tests are not valid.

Part B: Computer Exercises

1. The Capital Asset Pricing Model (CAPM), developed by John Lintner and William F. Sharpe in the 1960s, gives a convenient tool for assessing the performance of asset prices. According to the CAPM, when markets are in equilibrium, the riskiness of an asset relative to the riskiness of the entire asset market would be equal to the slope, β, in the relationship

$$\begin{aligned} & \text{(asset's excess return above the riskless rate)} \\ = \; & \beta \,\text{(excess return of a "market portfolio" above the riskless rate)} + u \end{aligned}$$

where a "market portfolio" is a portfolio containing every asset in the marketplace in proportion to its total value, and u is a mean zero, serially uncorrelated, homoskedastic disturbance term. The coefficient β, usually called the asset's "beta", measures the marginal contribution of the asset to a market portfolio's undiversifiable risk. If $\beta = 0.5$, then when the market excess return rises by 10%, this asset's excess return would rise by 5%.

In 1972, Black, Jensen and Scholes proposed that the validity of the CAPM can be tested by asking whether $\beta_0 = 0$ in

$$\text{(asset's excess return)} = \beta_0 + \beta_1 \,\text{(market's excess return)} + v$$

The data set CAPM2.DTA contains monthly observations for 16 years on the excess returns for six shares (two from each of three industries: the computer, paper, and airline industries). The excess returns for the market are given in $mreturn$ and excess returns for the six firms are given in $freturn$.[1] The first 192 observations pertain to firm 1, the next 192 pertain to firm 2, etc. The variable $firm$ identifies firms 1 to 6. For each of these 6 firms, test the null hypothesis that $\beta_0 = 0$. What do you conclude

[1] The excess return is calculated as the share's rate of return less the rate of return on a risk-free asset.

about the CAPM from these data? (You can use the `test` command to conduct hypothesis tests in Stata. Use the command: `help test` for more information.)

SOLUTION:

The Stata command for this question is `reg freturn mreturn if firm == i`, where `i` represents the firm number (i.e. `i=1,2,3,4,5,6`). You do not actually need to use the `test` command since the required hypothesis test is already calculated in Stata's regression output.

Firm 1:

```
      Source |       SS       df       MS              Number of obs =     192
-------------+------------------------------           F(  1,   190) =   41.62
       Model |  1.07325105     1  1.07325105           Prob > F      =  0.0000
    Residual |  4.89910446   190   .02578476           R-squared     =  0.1797
-------------+------------------------------           Adj R-squared =  0.1754
       Total |  5.97235551   191  .031268877           Root MSE      =  .16058

------------------------------------------------------------------------------
     freturn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     mreturn |   1.700561   .2635864     6.45   0.000     1.180629    2.220492
       _cons |  -.0043369   .0118146    -0.37   0.714    -.0276416    .0189677
------------------------------------------------------------------------------
```

Firm 2:

```
      Source |       SS       df       MS              Number of obs =     192
-------------+------------------------------           F(  1,   190) =   81.65
       Model |  .853008041     1  .853008041           Prob > F      =  0.0000
    Residual |  1.98495448   190  .010447129           R-squared     =  0.3006
-------------+------------------------------           Adj R-squared =  0.2969
       Total |  2.83796252   191  .014858443           Root MSE      =  .10221

------------------------------------------------------------------------------
     freturn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     mreturn |   1.516066   .1677799     9.04   0.000     1.185115    1.847016
       _cons |   .0105614   .0075203     1.40   0.162    -.0042727    .0253954
------------------------------------------------------------------------------
```

Firm 3:

```
      Source |       SS       df       MS              Number of obs =     192
-------------+------------------------------           F(  1,   190) =   98.40
       Model |  .511028451     1  .511028451           Prob > F      =  0.0000
    Residual |  .986709914   190   .00519321           R-squared     =  0.3412
-------------+------------------------------           Adj R-squared =  0.3377
       Total |  1.49773837   191  .007841562           Root MSE      =  .07206

------------------------------------------------------------------------------
     freturn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     mreturn |   1.173448   .1182931     9.92   0.000     .9401119    1.406785
       _cons |  -.0068681   .0053022    -1.30   0.197    -.0173269    .0035906
------------------------------------------------------------------------------
```

Firm 4:

```
      Source |       SS       df       MS              Number of obs =     192
-------------+------------------------------           F(  1,   190) =  152.06
       Model |  .577406391     1  .577406391           Prob > F      =  0.0000
    Residual |  .721486559   190  .003797298           R-squared     =  0.4445
-------------+------------------------------           Adj R-squared =  0.4416
       Total |  1.29889295   191  .006800487           Root MSE      =  .06162

------------------------------------------------------------------------------
     freturn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     mreturn |   1.247333    .101153    12.33   0.000     1.047805     1.44686
       _cons |   -.003961   .0045339    -0.87   0.383    -.0129043    .0049823
------------------------------------------------------------------------------
```

Firm 5:

```
      Source |       SS       df       MS              Number of obs =     192
-------------+------------------------------           F(  1,   190) =   25.55
       Model |   .17573607     1   .17573607           Prob > F      =  0.0000
    Residual |  1.30676742   190  .006877723           R-squared     =  0.1185
-------------+------------------------------           Adj R-squared =  0.1139
       Total |  1.48250349   191  .007761798           Root MSE      =  .08293

------------------------------------------------------------------------------
     freturn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     mreturn |   .6881323   .1361331     5.05   0.000     .4196058    .9566587
       _cons |    .010503   .0061018     1.72   0.087     -.001533     .022539
------------------------------------------------------------------------------
```

Firm 6:

```
      Source |       SS       df       MS              Number of obs =     191
-------------+------------------------------           F(  1,   189) =   64.74
       Model |  .098648566     1  .098648566           Prob > F      =  0.0000
    Residual |  .287987379   189  .001523743           R-squared     =  0.2551
-------------+------------------------------           Adj R-squared =  0.2512
       Total |  .386635945   190  .002034926           Root MSE      =  .03904

------------------------------------------------------------------------------
     freturn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     mreturn |   .5155932   .0640793     8.05   0.000     .3891907    .6419956
       _cons |   .0064084   .0028789     2.23   0.027     .0007295    .0120874
------------------------------------------------------------------------------
```

We cannot reject the null that $\beta_0 = 0$ at the 5% level of significance for 5 of the 6 cases. It would seem that the CAPM is a valid model for this sample of shares.

2. Nitrogen dioxide (NO_2) is a pollutant that attacks the human respiratory system; it increases the likelihood of respiratory illness. One common source of nitrogen dioxide is automobile exhaust. The file NO2POLLUTION.DTA contains a subset of 500 hourly observations made from October 2001 to August 2003. The variables in the data set are

lno2	Natural log of the concentration of NO_2 (particles)
lcars	Natural log of the number of cars per hour
temp	Temperature 2 metres above the ground (degrees C)
wndspd	Wind speed (metres/second)
tchng23	The temperature difference between 25 metres and 2 metres above ground (degrees C)
wnddir	Wind direction (degrees between 0 and 360)
hour	Hour of day
days	Day number from October 1, 2001

(a) Regress NO_2 concentration on the log of the number of cars, the two temperature variables, the two wind variables, and the time index (*days*). Which variables are significant at the 1% level? At the 5% level? At the 10% level? Interpret your results in full.

(b) Build a 95% confidence interval for the elasticity of NO_2 pollution with respect to car traffic and check that it matches the Stata output. Is NO_2 pollution elastic or inelastic with respect to car traffic?

(c) Test the hypothesis that, after controlling for *lcars, temp, tchng23* and *days*, the wind variables have no effect on NO_2 pollution.

(d) Does a temperature increase of 1 degree C have the same effect as a wind speed increase of 1 metre/second on NO_2 pollution?

(e) What is the estimated rate of change in NO_2 pollution per day?

(f) Is it correct to estimate the annual growth rate in NO_2 pollution by multiplying your estimate in (e) by 365? Briefly explain your answer.

(g) How much of the variation in the log of hourly levels of NO_2 pollution in this sample is accounted for by the variation in the regressors?

(h) How much of the variation in the log of hourly levels of NO_2 pollution in this sample could be accounted for by the variation in *days* alone?

SOLUTION:

(a)

```
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  6,   493) =   80.99
       Model |  139.553851     6  23.2589751           Prob > F      =  0.0000
    Residual |  141.580398   493  .287181335           R-squared     =  0.4964
-------------+------------------------------           Adj R-squared =  0.4903
       Total |  281.134249   499  .563395288           Root MSE      =  .53589

------------------------------------------------------------------------------
        lno2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lcars |   .4287319   .0230162    18.63   0.000       .38351    .4739538
        temp |  -.0246133   .0043664    -5.64   0.000    -.0331923   -.0160344
     tchng23 |   .1397861   .0259892     5.38   0.000     .0887229    .1908493
      wndspd |  -.1238996   .0142136    -8.72   0.000    -.1518262   -.0959729
      wnddir |   .0008034   .0003041     2.64   0.009     .0002059    .0014008
         day |   .0003261   .0001254     2.60   0.010     .0000797    .0005724
       _cons |   .8708752   .1793684     4.86   0.000     .5184544    1.223296
------------------------------------------------------------------------------
```

All variables except *day,* are significant at the 1% level. All variables are significant at the 5% level. All variables are significant at the 10% level. (Make sure you understand this!)

If the number of cars increases by 1%, hourly nitrogen dioxide concentration increases by approximately 0.4% on average (holding the other variables fixed).

If the temperature above the ground increases by 1 degree, hourly NO_2 concentration decreases by 2% $(100\,(-0.02))$ on average (holding the other variables fixed).

If the temperature difference increases by 1 degree, hourly NO_2 concentration increases by approximately 13% $(100\,(0.13))$ on average (holding the other variables fixed).

If the wind speed increases by 1 meter per second, hourly NO_2 concentration decreases by approximately 12% (100 (0.12)) on average (holding the other variables fixed).

If the wind direction changes by 1 degree, hourly NO_2 concentration increases by approximately 0.08% (100 (0.0008)) on average (holding the other variables fixed).

Over time (an extra day), hourly NO_2 concentration increases by approximately 0.03% (100 (0.0003)) on average (holding the other variables fixed).

Note that these are approximate changes in hourly NO_2 concentration, to compue the exact percentage change we must use the formula (page 190 of Wooldridge):

$$\%\Delta\widehat{y} = 100 \cdot \left[\exp\left(\widehat{\beta}_i \Delta x_i\right) - 1\right]$$

Calculate the exact percentage changes in NO_2 pollution for changes in two temperature variables, the two wind variables, and the time index (*days*). (The percentage in NO_2 pollution for a 1% change in the number of cars is exact because both variables are logged.)

(b) The confidence interval for β_{lcars} is given by: $\widehat{\beta}_{lcars} \pm c{\cdot}\mathrm{se}\left(\widehat{\beta}_{lcars}\right)$.

Degrees of freedom $= n - k - 1 = 500 - 6 - 1 = 493$. The 97.5^{th} percentile in a t_{493} distribution: $c = 1.96$.

Thus the confidence interval for β_{totwrk} is $0.4287319 - 1.96\,(0.0230162)$ or $[0.38362, 0.47384]$. This is very close to the Stata output.

This confidence interval means that if we obtained random samples repeatedly, the (unknown) population value β_{lcars} would lie in the interval $[0.38362, 0.47384]$ for 95% of the samples.

The range of likely values for the population parameter are less than one but greater than zero. This indicates that NO_2 pollution is relatively inelastic with respect to car traffic.

(c) $H_0 : \beta_{wnddir} = 0, \beta_{wndspd} = 0$

H_1 : At least one of β_{wnddir} or β_{wndspd} is different from zero.

Stata command: `test wnddir wndspd`

```
 ( 1)  wnddir = 0
 ( 2)  wndspd = 0

       F(  2,   493) =   47.93
            Prob > F =    0.0000
```

The test indicates that we can reject the null at the 1% significance level that the wind variables have no effect on NO_2 pollution.

You can also perform this test manually.

The unrestricted regression is the one from (a). The restricted regression is:

```
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  4,   495) =   81.98
       Model |  112.025946     4  28.0064864           Prob > F      =  0.0000
    Residual |  169.108303   495  .341632936           R-squared     =  0.3985
-------------+------------------------------           Adj R-squared =  0.3936
       Total |  281.134249   499  .563395288           Root MSE      =  .58449

------------------------------------------------------------------------------
        lno2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lcars |   .4264364   .0250441    17.03   0.000     .3772306    .4756422
        temp |  -.0235162   .0044167    -5.32   0.000    -.0321939   -.0148384
     tchng23 |    .191046   .0277649     6.88   0.000     .1364944    .2455977
         day |   .0003542   .0001354     2.62   0.009     .0000881    .0006203
       _cons |   .6060992   .1839351     3.30   0.001     .2447094     .967489
------------------------------------------------------------------------------
```

We can calculate the F statistic:

$$
\begin{aligned}
F &= \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} \\
&= \frac{(169.108303 - 141.580398)/2}{141.580398/(500-6-1)} \\
&= 47.927738
\end{aligned}
$$

With a 5% significance level, numerator $df = 2$, and denominator $df = 493$, the critical value is $c = 3.00$.

Since $F > c$, we can reject the null hypothesis and conclude that the wind variables do have an effect on NO_2 pollution.

(d) $H_0 : \beta_{temp} = \beta_{wndspd} \Rightarrow \beta_{temp} - \beta_{wndspd} = 0$
$H_1 : \beta_{temp} \neq \beta_{wndspd} \Rightarrow \beta_{temp} - \beta_{wndspd} \neq 0$
Stata command: `test temp=wndspd`

```
( 1)  temp - wnddir = 0

       F(  1,   493) =   32.09
            Prob > F =    0.0000
```

We can reject the null at the 1% significance level, and conclude that a temperature increase of 1 degree C does not have the same effect as a wind speed increase of 1 metre/second on NO_2 pollution.

You can also perform this test manually.

Use the Stata command `correlate, _coef covariance` immediately after the regression command to obtain the variance–covariance matrix for the estimated coefficients:

```
             |    lcars     temp  tchng23   wndspd   wnddir      day    _cons
-------------+---------------------------------------------------------------
       lcars |   .00053
        temp | -8.8e-06  .000019
     tchng23 |  .000112  .000035  .000675
      wndspd | -.000013 -9.1e-06  .000066  .000202
      wnddir | -4.4e-07 -4.9e-07 -4.5e-07  8.6e-07  9.2e-08
         day |  1.0e-07 -8.2e-08  6.5e-07  1.5e-07  4.8e-09  1.6e-08
       _cons | -.003633  .000163 -.001248 -.000697 -.000014 -6.8e-06  .032173
```

Now you can compute the test statistic:

$$t = \frac{\widehat{\beta}_{temp} - \widehat{\beta}_{wndspd}}{\text{se}\left(\widehat{\beta}_{temp} - \widehat{\beta}_{wndspd}\right)}$$

where

$$\text{se}\left(\widehat{\beta}_{temp} - \widehat{\beta}_{wndspd}\right) = \left\{\left[\text{se}\left(\widehat{\beta}_{temp}\right)\right]^2 + \left[\text{se}\left(\widehat{\beta}_{wndspd}\right)\right]^2 - 2s_{\widehat{\beta}_{temp},\widehat{\beta}_{wndspd}}\right\}^{1/2}$$

Therefore

$$\begin{aligned} t &= \frac{-.0246133 - (-.1238996)}{\left[(.0043664)^2 + (.0142136)^2 + 2\left(-9.1 \times 10^{-06}\right)\right]^{1/2}} \\ &= \frac{0.099286}{0.014244} \\ &= 6.9704 \end{aligned}$$

The 95% critical value is 1.96, so we can reject the null hypothesis.

(e)

$$100 \cdot (\exp(.0003261) - 1) \times 24$$

Thus $N0_2$ concentration increases by 0.78276% for every additional day given the way the data has been sampled.

(f) No this would be incorrect since every day has not been sampled, we have only observed 500 hours over 3 years. If we did have daily data, then we would find the yearly change not by multiplying the answer in (e) by 365, but by multiplying the coefficient by 365 and then exponentiating it.

(g) The R^2 indicates that about 49% of the variation in the log of hourly levels of NO_2 pollution in this sample is accounted for by the variation in the regressors.

(h) We regress *lno2* on *days*:

```
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =    0.02
       Model |  .011610829     1  .011610829           Prob > F      =  0.8860
    Residual |  281.122638   498  .564503289           R-squared     =  0.0000
-------------+------------------------------           Adj R-squared = -0.0020
       Total |  281.134249   499  .563395288           Root MSE      =  .75133

------------------------------------------------------------------------------
        lno2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         day |    .000024   .0001674     0.14   0.886    -.0003048    .0003528
       _cons |   3.690916   .0618768    59.65   0.000     3.569344    3.812488
------------------------------------------------------------------------------
```

The R^2 indicates that none of the variation of the log of hourly levels of NO_2 pollution is explained by the variation in *days* alone.

3. Consider a model where the return to education depends upon the amount of work experience (and vice versa):

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 educ \cdot exper + u.$$

where *wage* denotes monthly earnings, *educ* denotes years of education and *exper* denotes years of work experience.

(a) Show that the return to another year of education (in decimal form), holding *exper* fixed, is $\beta_1 + \beta_3 exper$.

(b) State the null hypothesis that the return to education does not depend on the level of *exper*. What do you think is the appropriate alternative?

(c) Use the data in WAGE2.DTA to test the null hypothesis in (b) against your stated alternative. (In order to estimate the regression model, you will first need to create a new variable: `gen educXexper = educ*exper` and then incorporate this interaction term into the regression: `reg lwage educ exper educXexper`)

(d) Let θ_1 denote the return to education (in decimal form), when $exper = 10$: $\theta_1 = \beta_1 + 10\beta_3$. Obtain $\widehat{\theta}_1$ and a 95% confidence interval for $\widehat{\theta}_1$. (*Hint*: Write $\beta_1 = \theta_1 - 10\beta_3$ and plug this into the equation; then rearrange. This gives the regression for obtaining the confidence interval for θ_1.)

SOLUTION:

(a) Holding *exper* (and the elements in u) fixed, we have

$$\begin{aligned}\Delta \log(wage) &= \beta_1 \Delta educ + \beta_3 (\Delta educ)\, exper \\ &= (\beta_1 + \beta_3 exper)\, \Delta educ\end{aligned}$$

or

$$\frac{\Delta \log(wage)}{\Delta educ} = (\beta_1 + \beta_3 exper)$$

This is the approximate proportionate change in *wage* given one more year of education.

(b) $H_0 : \beta_3 = 0$. If we think that education and experience interact positively – so that people with more experience are more productive when given another year of education – then $\beta_3 > 0$ is the appropriate alternative.

(c) The estimated equation is

$$\begin{aligned}\widehat{\log(wage)} &= \underset{(0.24)}{5.95} + \underset{(.0174)}{.0440}\ educ - \underset{(.0200)}{.0215}\ exper + \underset{(.00153)}{.00320}\ educ \cdot exper \\ n &= 935, \quad R^2 = .135, \quad \overline{R}^2 = .132\end{aligned}$$

The t statistic on the interaction term is about 2.13,which gives a p-value below .02 against $H_1 : \beta_3 > 0$. Therefore, we reject $H_0 : \beta_3 = 0$ against $H_1 : \beta_3 > 0$ at the 2% level.

(d) We rewrite the equation as

$$\log(wage) = \beta_0 + \theta_1 educ + \beta_2 exper + \beta_3 educ\,(exper - 10) + u.$$

and run the regression $\log(wage)$ on *educ*, *exper*, and $educ(exper - 10)$. We want the coefficient on *educ*. We obtain $\widehat{\theta}_1 \approx 0761$ and $se(\widehat{\theta}_1) \approx .0066$. The 95% CI for θ_1 is about .063 to .089.