Econometrics - TA session 1

QUESTION 1

Assume a regression model with intercept. Data are as follows:

$$X'X = \begin{bmatrix} 33 & 0 & 0 \\ 0 & 40 & 20 \\ 0 & 20 & 60 \end{bmatrix}, \; X'y = \begin{bmatrix} 132 \\ 24 \\ 92 \end{bmatrix}, \; \sum (y_i - \bar{y})^2 = 150$$

(i) What is the sample size?

(ii) Compute the regression equation with using LS.

(iii) Estimate the standard error of b_2 and test the hypothesis that β_2 is zero. (Write down the null and the alternative. Derive the test statistic and explain how you will decide for the null or for the alternative hypothesis.)

(iv) Test the hypothesis that $\beta_2 + \beta_3 = 1$. (Write down the null and the alternative. Derive the test statistic and explain how you will decide for the null or for the alternative hypothesis.)

(v) Compute R^2.

<u>Solution</u>

(i) Sample size: The sample size is $n = 33$. To see this, notice that since we have an intercept in the model

$$
\begin{aligned}
X'X &= \begin{bmatrix} 1 & \cdots & 1 \\ x_{12} & \cdots & x_{n2} \\ x_{13} & \cdots & x_{n3} \end{bmatrix} \begin{bmatrix} 1 & x_{12} & x_{13} \\ \vdots & \vdots & \vdots \\ 1 & x_{n2} & x_{n3} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^{n} 1 & \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} x_{i3} \\ \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} x_{i2}^2 & \sum_{i=1}^{n} x_{i2}x_{i3} \\ \sum_{i=1}^{n} x_{i3} & \sum_{i=1}^{n} x_{i2}x_{i3} & \sum_{i=1}^{n} x_{i3}^2 \end{bmatrix} \\
&= \begin{bmatrix} n & n\bar{x}_2 & n\bar{x}_3 \\ n\bar{x}_2 & \sum_{i=1}^{n} x_{i2}^2 & \sum_{i=1}^{n} x_{i2}x_{i3} \\ n\bar{x}_3 & \sum_{i=1}^{n} x_{i2}x_{i3} & \sum_{i=1}^{n} x_{i3}^2 \end{bmatrix} \\
&= \begin{bmatrix} 33 & 0 & 0 \\ 0 & 40 & 20 \\ 0 & 20 & 60 \end{bmatrix}
\end{aligned}
$$

(ii) Compute the regression equation for LS: We know that

$$
\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = [X'X]^{-1}X'y = \begin{bmatrix} 33 & 0 & 0 \\ 0 & 40 & 20 \\ 0 & 20 & 60 \end{bmatrix}^{-1} \begin{bmatrix} 132 \\ 24 \\ 92 \end{bmatrix}
$$

Now to find the inverse $[X'X]^{-1}$, we use the rule for inverse of block diagonal matrices

$$
\begin{aligned}
\begin{bmatrix} 33 & 0 & 0 \\ 0 & 40 & 20 \\ 0 & 20 & 60 \end{bmatrix}^{-1} &= \begin{bmatrix} (33)^{-1} & \mathbf{0} \\ \mathbf{0} & \begin{bmatrix} 40 & 20 \\ 20 & 60 \end{bmatrix}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{33} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2000} \begin{bmatrix} 60 & -20 \\ -20 & 40 \end{bmatrix} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{33} & 0 & 0 \\ 0 & \frac{3}{100} & -\frac{1}{100} \\ 0 & -\frac{1}{100} & \frac{2}{100} \end{bmatrix}
\end{aligned}
$$

So,

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{33} & 0 & 0 \\ 0 & \frac{3}{100} & -\frac{1}{100} \\ 0 & -\frac{1}{100} & \frac{2}{100} \end{bmatrix} \begin{bmatrix} 132 \\ 24 \\ 92 \end{bmatrix} = \begin{bmatrix} 4 \\ -0.2 \\ 1.6 \end{bmatrix}$$

So, the estimated regression is

$$y_i = 4 - 0.2x_{i2} + 1.6x_{i3} + e_i.$$

(iii) Standard error of b_2 and test hypothesis that β_2 is zero.

We know that

$$se(b_2) = \sqrt{\hat{V}(b_2 \mid X)}$$

and that

$$\begin{aligned} V(b \mid X) = \sigma^2(X'X)^{-1} &= \begin{bmatrix} V(b_1 \mid X) & \text{Cov}\,(b_1, b_2 \mid X) & \text{Cov}\,(b_1, b_3 \mid X) \\ \text{Cov}\,(b_1, b_2 \mid X) & V(b_2 \mid X) & \text{Cov}\,(b_2, b_3 \mid X) \\ \text{Cov}\,(b_1, b_3 \mid X) & \text{Cov}\,(b_2, b_3 \mid X) & V(b_3 \mid X) \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} 0.03 & 0 & 0 \\ 0 & 0.03 & -0.01 \\ 0 & -0.01 & 0.02 \end{bmatrix} \\ \Rightarrow V(b_2 \mid X) &= 0.03\sigma^2 \end{aligned}$$

To obtain $\hat{V}(b_2 \mid X)$ we replace σ^2 with s^2, where

$$s^2 = \frac{e'e}{n-k} = \frac{e'e}{30}.$$

The normal equations imply the orthogonal decomposition

$$\begin{aligned} \sum_{i=1}^{n} y_i^2 = y'y = \sum_{i=1}^{n} \hat{y}_i^2 + \sum_{i=1}^{n} e_i^2 = \hat{y}'\hat{y} + e'e \\ \Rightarrow \qquad e'e = y'y - \hat{y}'\hat{y} \end{aligned}$$

We are told that

$$\begin{aligned} \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = 150 \\ \Rightarrow y'y = \sum_{i=1}^{n} y_i^2 = 150 + n\bar{y}^2 \end{aligned}$$

We know $n = 33$ and so we need to find $\bar{y}$. We know

$$X'y = \begin{bmatrix} 1 & \cdots & 1 \\ x_{12} & \cdots & x_{n2} \\ x_{13} & \cdots & x_{n3} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i2} y_i \\ \sum_{i=1}^{n} x_{i3} y_i \end{bmatrix} = \begin{bmatrix} 132 \\ 24 \\ 92 \end{bmatrix}.$$

Therefore,

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{132}{33} = 4.$$

Going back

$$y'y = 150 + 33 \times 4^2 = 678.$$

Now need to find $\hat{y}'\hat{y}$, where $\hat{y} = Py = X(X'X)^{-1}X'y$.

$$\begin{aligned} \hat{y}'\hat{y} &= y'P'Py \\ &= y'Py \\ &= y'X \underbrace{(X'X)^{-1}X'y}_{=b} \\ &= [132 \quad 24 \quad 92] \begin{bmatrix} 4 \\ -0.2 \\ 1.6 \end{bmatrix} \\ &= 670.4. \end{aligned}$$

Consequently,

$$e'e = y'y - \hat{y}'\hat{y} = 678 - 670.4 = 7.6$$

Finally,

$$s^2 = \frac{e'e}{30} = 0.253$$

Hypothesis testing: We have the following two hypotheses:

$$H_0: \beta_2 = \beta_2^0 = 0, \text{ vs } H_1: \beta_2 \neq 0.$$

To test this hypothesis we employ the test statistic

$$t = \frac{b_2 - \beta_2^0}{se(b_2 - \beta_2^0)} = \frac{b_2 - \beta_2^0}{se(b_2)},$$

which under the assumption $\varepsilon \mid X \sim N(0, \sigma^2 I_n)$ and the null hypothesis $t \sim t_{n-k} = t_{30}$. At a size of test $\alpha = 0.05$, we are going to reject the null if $|t| > |t_{0.025,30}| \approx 2.042$, where $\Pr(t > 2.042) = 0.025 = \Pr(t < -2.042)$.

So, in our case

$$t = \frac{b_2 - \beta_2^0}{se(b_2)} = \frac{-0.2 - 0}{\sqrt{0.253 \times 0.03}} = -2.29$$

So, the null is rejected.

(iv) Test the hypothesis that $\beta_2 + \beta_3 = 1$

Two approaches:

t-test: The hypotheses is

$$H_0: \beta_2 + \beta_3 = 1, \text{vs } H_1: \beta_2 + \beta_3 \neq 1$$

the test statistic would be

$$t = \frac{b_2 + b_3 - 1}{\text{se }(b_2 + b_3)} = \frac{b_2 + b_3 - 1}{\sqrt{\hat{\text{Var}}(b_2) + \hat{\text{Var}}(b_2) + 2\hat{\text{Cov}}(b_2, b_3)}}$$

which again under normality assumption and the null, we have that $t \sim t_{30}$. We reject the null at size of test 5% if $|t| > t_{0.025,30} \approx 2.042$, where $\Pr(t > 2.042) = 0.025 = \Pr(t < -2.042)$.

In our example,

$$\begin{aligned} t &= \frac{b_2 + b_3 - 1}{\sqrt{\text{Var }(b_2) + \hat{\text{Var}}(b_2) + 2\text{Cov }(\hat{b}_2, b_3)}} \\ &= \frac{-0.2 + 1.6 - 1}{\sqrt{0.253(0.03 + 0.02 + 2(-0.01))}} \\ &= \frac{0.4}{\sqrt{0.253(0.03)}} \\ &= \sqrt{21.08} \\ &= 4.59 \end{aligned}$$

So, we reject the null. F-test: The hypotheses are

$$H_0: R\beta = r, \text{vs } H_1: R\beta \neq r,$$

where

$$R\beta = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = 1 = r.$$

The test statistic would be

$$F = \frac{(Rb - r)'(s^2 R(X'X)^{-1}R')^{-1}(Rb - r)}{1},$$

which again under normality assumption and the null, we have that $F \sim F(1,30)$. We reject the null at size of test 5% if $F > F_{0.05,(1,30)} \approx 4.17$, where $\Pr(F > 4.17) = 0.05$.

In our example,

$$\begin{aligned} F &= \frac{(Rb - r)'(s^2 R(X'X)^{-1}R')^{-1}(Rb - r)}{1} \\ &= \frac{(b_2 + b_3 - 1)\left([0 \quad 1 \quad 1]\begin{bmatrix} 0.03 & 0 & 0 \\ 0 & 0.03 & -0.01 \\ 0 & -0.01 & 0.02 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right)^{-1}(b_2 + b_3 - 1)}{s^2} \\ &= \frac{(-0.2 + 1.6 - 1)\left([0 \quad 0.03 - 0.01 \quad -0.01 + 0.02]\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right)^{s^2}(-0.2 + 1.6 - 1)}{0.253} \\ &= \frac{0.4(0.03)^{-1}0.4}{0.253} \\ &= 21.08. \end{aligned}$$

So, we reject the null.

Remark: Since one restriction we have the result that $t^2 = F$. In the case of one linear restriction the t-test should be preferred to F-test.

(v) Compute R^2

Centered R^2 :

$$R_c^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{7.6}{150} = 0.949.$$

Uncentered R^2 :

$$R_u^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} y_i^2} = 1 - \frac{7.6}{678} = 0.9887$$

QUESTION 2

Consider the linear regression model

$$y = X\beta + u,\ u \sim N(0, \sigma^2 I_N).$$

where $y = N \times 1, X = N \times K$. You know that (true) $\sigma^2 = 2$ and that

$$X'X = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}.$$

In a sample of 32 observations, the LS coefficients are $b_1 = 2, b_2 = 2$.

1 Test at the 5% significance level the joint null hypothesis that $\beta_1 = \beta_2 = 3$. State analytically the alternative hypothesis against which you are testing.

2 Assume that you do not know σ^2 but you are given that its estimate is $s^2 = 2$. In what sense, can the statistic you used to test the joint hypothesis be valid?

It is suggested that you find critical values using proper tables and p-values with the software of your preference using proper commands.

Solution

1 The null hypothesis

$$H_0: \beta_1 = \beta_2 = 3$$

can be written as $H_0: R\beta = r$ with

$$R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \; r = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \; \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}.$$

For the Normal Classical Linear Regression model, we have that (conditional on X) $b \sim N(\beta, \sigma^2 (X'X)^{-1})$. It follows that, under $H_0: R\beta = r$,

$$\begin{aligned} & Rb \sim N(r, \sigma^2 R(X'X)^{-1} R') \\ \Rightarrow \chi \equiv \quad & (Rb - r)' [\sigma^2 R(X'X)^{-1} R']^{-1} (Rb - r) \sim \chi^2_{(2)}. \end{aligned}$$

Since σ^2 is known, χ can be computed for a given sample and its (exact) sampling distribution can be used to conduct hypothesis testing.

Given the particular form of R in this case, the statistic takes the form

$$\chi = \frac{(b-r)'(X'X)(b-r)}{\sigma^2}$$

which is convenient since it does not require inversion of $(X'X)$. With our data, $(b - r)' = (-1 \quad -1)$ and

$$\chi = \frac{1}{2} (-1 \quad -1) \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{pmatrix} -1 \\ -1 \end{pmatrix} = 3.50$$

From the table of $\chi^2_{(2)}$ we get that the $\chi^2_{(2),0.05} = 5.99$ and the null hypothesis is not rejected at level of statistical significance $\alpha = 5\%$.

2 The test statistic χ is distributed as chi-sq with 2 degrees of freedom for any sample size n. When the σ^2 is replaced by the unbiased estimator s^2 the statistic, say W, is validated only asymptotically:

$$W = (Rb - r)'[s^2R(X'X)^{-1}R']^{-1}(Rb - r) \xrightarrow{d} \chi^2_{(2)}.$$

Since the sample size ($n = 32$) is relatively small, but we maintain the normality assumption, a natural choice is to use the F statistic for testing this null hypothesis:

$$F = \frac{W}{2} = 1.75$$

with critical value $F_{(2,30),0.05} = 3.3158$.

Question 3

Suppose that x follows the Weibull distribution

$$f(x) = \alpha\beta x^{\beta-1}e^{\alpha x^{\beta}}, x \geq 0, \alpha, \beta > 0$$

a. Obtain the log-likelihood function for a random sample of n observations.

b. Obtain the likelihood equations for maximum likelihood estimation of α and β. Note that the first provides an explicit solution for α in terms of the data and β. But, after inserting this in the second, we obtain only an implicit solution for β. How would you obtain the maximum likelihood estimators?

c. Obtain the second derivatives matrix of the log-likelihood with respect to α and β. The exact expectations of the elements involving β involve the derivatives of the gamma function and are quite messy analytically. Of course, your exact result provides an empirical estimator. How would you estimate the asymptotic covariance matrix for your estimators in part b?

d. Prove that $\alpha\beta Cov[\ln x, x^{\beta}] = 1$. (Hint: The expected first derivatives of the log-likelihood function are zero.)

Solution

a. We form the likelihood function:

$$L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta) = \alpha^n\beta^n\left(\prod_{i=1}^{n} x_i\right)^{\beta-1} e^{-a\sum x_i^{\beta}}$$

It is usually simpler to work with the log of the likelihood function:

$$\ln\big(L(\theta|x)\big) = nlna + nln\beta + (\beta - 1)\sum_{\iota=1}^{n} lnx_i - a\sum_{i=1}^{n} x_i^{\beta}$$

b. The F.O.C are given by:

$$\frac{\partial lnL}{\partial a} = \frac{n}{a} - \sum x_i^\beta = 0$$

$$\frac{\partial lnL}{\partial \beta} = \frac{n}{\beta} + \sum lnx_i - \alpha \sum lnx_i x_i^\beta = 0$$

Since the first likelihood equation implies that at the maximum, $\hat{\alpha} = n / \sum x_i^\beta$ one approach to obtain the ml estimators would be to scan over the range of β and compute the implied value of α. Two practical complications are the allowable range of β and the starting values to use for the search.

c. The second derivatives are:

$$\frac{\partial^2 lnL}{\partial a^2} = -\frac{n}{a^2}$$

$$\frac{\partial^2 lnL}{\partial \beta^2} = -\frac{n}{\beta^2} - \alpha \sum (lnx_i)^2 x_i^\beta$$

$$\frac{\partial^2 lnL}{\partial \alpha \beta} = -\sum lnx_i x_i^\beta$$

If we had estimates in hand, the simplest way to estimate the expected values of the Hessian would be to evaluate the expressions above at the maximum likelihood estimates, then compute the negative inverse.

d. First, the expected value of the first FOC is zero which implies the following:

$$E\left(\frac{\partial lnL}{\partial a}\right) = 0 \Leftrightarrow \frac{n}{a} - \sum E(x_i^\beta) = 0 \Leftrightarrow \frac{n}{a} - nE(x_i^\beta) = 0 \Leftrightarrow E(x_i^\beta) = 1/a$$

Now, the expected value of the second FOC is also zero:

$$E\left(\frac{\partial lnL}{\partial \beta}\right) = 0 \Leftrightarrow \frac{n}{\beta} + E\left(\sum lnx_i\right) - \alpha E\left(\sum lnx_i x_i^\beta\right) = 0$$

Divide by n, use the fact that every term in the sum has the same expectation and substitute $a = 1/E(x_i^\beta)$ to obtain:

$$\frac{1}{\beta} + E(lnx_i) - \alpha E\left(lnx_i x_i^\beta\right) = 0$$

Now multiply by $E(x_i^\beta) = \frac{1}{a}$ to obtain

$$\frac{1}{a\beta} + E(lnx_i)E\left(x_i^\beta\right) - E\left(lnx_i x_i^\beta\right) = 0 \Leftrightarrow cov(lnx_i, x_i) = \frac{1}{a\beta}$$

Since $cov(lnx_i, x_i) \equiv E\left(lnx_i x_i^\beta\right) - E(lnx_i)E\left(x_i^\beta\right)$.