Econ616
Homework #4
Answer key

Question 1

Model $Y=\alpha_1+\gamma_2 E_2+\gamma_3 E_3+u$

Solution

A possible solution migth be:

Define $X=[i\ E_2\ E_3]$ and thus

$$X'X=\begin{bmatrix} i'i & i'E_2 & i'E_3 \\ E_2'i & E_2'E_2 & E_2'E_3 \\ E_3'i & E_3'E_2 & E_3'E_3 \end{bmatrix}=\begin{bmatrix} N & n_2 & n_3 \\ n_2 & n_2 & 0 \\ n_3 & 0 & n_3 \end{bmatrix}$$

where $N=n_1+n_2+n_3$ and n_i is a number of observations in class i. Then

$$(X'X)^{-1}=\begin{bmatrix} \frac{-1}{-N+n_2+n_3} & \frac{1}{-N+n_2+n_3} & \frac{1}{-N+n_2+n_3} \\ \frac{1}{-N+n_2+n_3} & \frac{-N+n_3}{n_2(-N+n_2+n_3)} & \frac{-1}{-N+n_2+n_3} \\ \frac{1}{-N+n_2+n_3} & \frac{-1}{-N+n_2+n_3} & \frac{-N+n_2}{n_3(-N+n_2+n_3)} \end{bmatrix} \text{ and}$$

$$X'y=\begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n_2} Y_{2i} \\ \sum_{i=1}^{n_3} Y_{3i} \end{bmatrix}$$

From that follows

$$\begin{bmatrix} a_1 \\ c_2 \\ c_3 \end{bmatrix}=(X'X)^{-1}X'Y=\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2-\bar{Y}_1 \\ \bar{Y}_3-\bar{Y}_1 \end{bmatrix}.$$

However there is <u>more elegant solution</u>: $i=E_1+E_2+E_3$. Thus

$$Y=\alpha_1 i+\gamma_2E_2+\gamma_3E_3+u=\alpha_1(E_1+E_2+E_3)+\gamma_2E_2+\gamma_3E_3+u=\alpha_1E_1+(\alpha_1+\gamma_2)E_2+(\alpha_1+\gamma_3)E_3+u$$

Since binary variables are mutually exclusive and exhaustive, they are orthogonal, thus one may run 3 separate regressions to estimate the coefficients:

$$\begin{bmatrix} a_1 \\ a_1+c_2 \\ a_1+c_3 \end{bmatrix} = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \bar{Y}_3 \end{bmatrix}.$$ The rest follows…

Question 2

Solution

Using a software package one will obtain

$$Y=9E_1+13E_2+22.5E_3-2S_2+e$$

And

$$Y=22.5-13.5E_1-9.5E_2-2S_2+e$$

Alternatively one may recall that binary variables are mutually exclusive and exhaustive and thus their sum is identically equal zero and work out the paramater estimates from the given regression output without actually rerunning the regression. (see Question 1)

Question 3

Solution

The regression estimated is

$$Y=10+3E_2+11E_3-4.5S_2+2.5E_2S_2+5.5E_3S_2+e$$

Lets regress Y on E_1 E_2 E_3 and S_1. There are 6 possible interactions terms that may be included into the model, however of them should be disregarded due to orthogonality of binary variables – E_1E_2, E_1E_3, E_2E_3. Next out of remaining 3 interactions terms should be excluded because these interactions terms and the 4 binary variables are not linearly independent. For example,

$$E_1S_1=(i-E_2-E_3)S_1=S_1-E_2S_1-E_3S_1.$$

Suppose we dropped E_1S_1 then the estimated regression is

$Y=5.5E_1+11E_2+22E_3+4.5S_1-2.5E_2S_1-5.5E_3S_1+e$

And it leads to the mean incomes give in the table 4.6.

Question 4

Check your notes on Chow test or Greene A test based on loss of fit p. 282 and on …

In general $e'e=y'y-b'X'y=y'y-y'X(X'X)^{-1}X'y$

Unrestricted models

Rural $y_R=X_R\beta_R+\varepsilon_1$

$$e_1'e_1 = 30 - \begin{bmatrix}10 & 20\end{bmatrix}\begin{bmatrix}20 & 20\\20 & 25\end{bmatrix}^{-1}\begin{bmatrix}10\\20\end{bmatrix} = 5$$

Urban $y_U=X_U\beta_U+\varepsilon_2$

$$e_2'e_2 = 24 - \begin{bmatrix}8 & 20\end{bmatrix}\begin{bmatrix}10 & 10\\10 & 20\end{bmatrix}^{-1}\begin{bmatrix}8\\20\end{bmatrix} = 3.2$$

Restricted model

Rural and Urban $\begin{bmatrix}y_R\\y_U\end{bmatrix} = \begin{bmatrix}X_R\\X_U\end{bmatrix}\beta + \begin{bmatrix}\varepsilon_1\\\varepsilon_2\end{bmatrix}$

$X^{*\prime}=[X_R' \; X_U']$ and thus $X^{*\prime}X^*=X_R'X_R+X_U'X_U$

$$X^{*\prime}X^* = \begin{bmatrix}30 & 30\\30 & 45\end{bmatrix}$$

$y^{*\prime}=[y_R' \; y_U']$ and thus $X^{*\prime}y^*= X_R'y_R+X_U'y_U$

$$X^{*\prime}y^* = \begin{bmatrix}18\\40\end{bmatrix}$$

$$e^{*\prime}e^* = 54 - \begin{bmatrix}18 & 40\end{bmatrix}\begin{bmatrix}30 & 30\\30 & 45\end{bmatrix}^{-1}\begin{bmatrix}18\\40\end{bmatrix} = 10.93$$

Solution

Chow Test (F Test)

H_0: $\beta_R=\beta_U$
H_A: $\beta_R<>\beta_U$

$$F=\frac{(RRSS-URSS)/k}{URSS/(n-2k)}=\frac{(e^{*\prime}e^{*}-e_1{}'e_1-e_2{}'e_2)/k}{(e_1{}'e_1+e_2{}'e_2)/(n-2k)}=\frac{(10.93-5-3.2)/2}{(5+3.2)/(30-2*2)}=4.3$$

5% critical value is 3.37 – reject the null hypothesis.

Question 5

Model $S=\alpha+\beta_1ED+\beta_2IQ+\beta_3EX+\beta_4Sex+\beta_5DF+\beta_6DE+\varepsilon$

The reference group consists of bilingual.

(a) $E(S|Sex=1)-E(S|Sex=0)=\beta_4$. Check the sign and statistical significance of β_4, contruct a t test statistic.
(b) $E(S|DF=0;DE=0)-E(S|DF=1;DE=0)=-\beta_6$
$E(S|DF=0;DE=0)-E(S|DF=0;DE=1)=-\beta_5$
(c) H_0: $\beta_5=\beta_6$
H_A: $\beta_5<>\beta_6$
Under H_0 say $\beta_5=\beta_6=\beta$ then
$S=\alpha+\beta_1ED+\beta_2IQ+\beta_3EX+\beta_4Sex+\beta(DF+DE)+\varepsilon$
Then one may run the usual F test…
(d) $E(S|DF=1;Sex=1)-E(S|DE=1;Sex=0)=\beta_5+\beta_4--\beta_6$
H_0: $\beta_5+\beta_4--\beta_6=0$
H_A: $\beta_5+\beta_4--\beta_6<>0$
Then one may run the usual F test…
(e) Introduce an interaction term Sex*Ex into the model and test the significance of the interaction term coefficient.

Question 6

Model 1 $y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{2i}^2+u_i$

Model 2 $y_i=\beta_1^*+\beta_2^*X_{2i}+v_i$

Since $v_i=\beta_3X_{2i}^2+u_i$

b_2 is a biased estimator

$$E(b_2)=\beta_2+\beta_3\frac{\sum_{i=1}^{n}(x_{2i})X_{2i}^2}{\sum_{i=1}^{n}x_{2i}^2}$$

where x_{2i} is a mean devation form of X_{2i}. Note that a numerator of the fraction reminds a sample covariance and thus the whole fraction may be regarded as a correlation coefficient of X_2 and X_2^2.

Solution

In general, the direction of bias is given the sign of β_3 and the correlation coefficient:

	$\rho(X_2;X_2^2)>0$	$\rho(X_2;X_2^2)<0$
$\beta_3>0$	$E(b_2)-\beta_2>0$	$E(b_2)-\beta_2<0$
$\beta_3<0$	$E(b_2)-\beta_2<0$	$E(b_2)-\beta_2>0$

In this particular case you may go further since you may tell more about the correlation coefficient:

	$X_{2i}^2>0$
$x_{2i}>0,\ \forall i$	$\rho(X_2;X_2^2)>0$
$x_{2i}<0,\ \forall i$	$\rho(X_2;X_2^2)<0$

<u>Question 7</u>

Model 1 $y=\alpha_1D_1+\alpha_2D_2+\alpha_3D_3+\alpha_4D_4+\varepsilon$

Model 2 $y=\alpha+\alpha_2D_2+\alpha_3D_3+\alpha_4D_4+\varepsilon$

Since binary variables are mutually exclusive and exhaustive $D_1+D_2+D_3+D_4=i$.

Solution

Lets deal with the first model for now. Define $D=[D_1\ D_2\ D_3\ D_4]$ thus

$$D'D=\begin{bmatrix}n&0&0&0\\0&n&0&0\\0&0&n&0\\0&0&0&n\end{bmatrix}\text{ and }(D'D)^{-1}=\begin{bmatrix}1/n&0&0&0\\0&1/n&0&0\\0&0&1/n&0\\0&0&0&1/n\end{bmatrix}\text{ and }D'y=\begin{bmatrix}\sum_{i=1}^{n}y_{1i}\\\sum_{i=1}^{n}y_{2i}\\\sum_{i=1}^{n}y_{3i}\\\sum_{i=1}^{n}y_{4i}\end{bmatrix}$$

where n is a number of observations in each quarter.

OLS gives $d_1 = (D'D)^{-1}D'y = \begin{bmatrix} \bar{y}_1 & \bar{y}_2 & \bar{y}_3 & \bar{y}_4 \end{bmatrix}$ - quite an expected result given orthogonality of binary variables.

Now lets deal with the second model, define D*=DP where P is

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \text{ and thus } d_2 = (D^{*\prime}D^*)^{-1}D^{*\prime}y = P^{-1}(D'D)^{-1}(P')^{-1}P'y = P^{-1}d_1$$

It follows that $d_2 = \begin{bmatrix} \bar{y}_1 & \bar{y}_2 - \bar{y}_1 & \bar{y}_3 - \bar{y}_1 & \bar{y}_4 - \bar{y}_1 \end{bmatrix}$ since

$$P^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}$$

. Reminds Question 1, doesn't it?

Now lets introduce x into both models.

Model 1 $y=\alpha_1 D_1+\alpha_2 D_2+\alpha_3 D_3+\alpha_4 D_4+\beta_1 x+\varepsilon$

Model 2 $y=\alpha+\alpha_2 D_2+\alpha_3 D_3+\alpha_4 D_4+\beta_2 x+\varepsilon$

Recall the results of partitioned regression:

$b_1 = (x^{*\prime}x^*)^{-1}x^{*\prime}y$ where $x^*=x-D(D'D)^{-1}D'x$

and

$b_2 = (x^{\#\prime}x^{\#})^{-1}x^{\#\prime}y$ where $x^{\#}=x-D^*(D^{*\prime}D^*)\text{-}1D^{*\prime}x=x-(DP)P^{-1}(D'D)^{-1}P'^{-1}(DP)'x=x^*$

Thus $b_1=b_2$.

Question 8

Model $\ln(price)=6.4+0.056H+0.249W+0.23L+0.10V+0.23T+0.090A+0.088P+0.109B+0.157C-0.044D_1-0.015D_2+0.019D_3+0.044D_5+0.023D_6+e$

Solution

(a) plug in the benchmark numbers and zeroes for binary variables to get a predicted ln(price)=7.877 and thus a predicted price=2637.82.

(b) $\partial \ln(\text{price}) / \partial X_i$ where Xi is any of the independent variables of the model gives the percentage change in price. It would cost 5.4% more to buy a car with the V8 (-4.4%) and power steering (+8.8%).

(c) The price increased by D_3–D_2=1.9%--(--1.5%)=3.4% from 1956 to 1957.

(d) Then coefficient of T equal to 2.3% overstates the additional cost of having hardtop. The weight of the car becomes a function of T: W_T=W+0.1T. Thus:

ln(price)=…+0.249(W+0.1T)+…+0.23T+…+e

It follows that $\partial \ln(\text{price}) / \partial T$ =b_T+0.1*0.249=0.023 and b_T=--0.0019.

(e) H_0: six binary variable coeffiecients are zeroes

H_A: at least one coefficient is not zero

$$F = \frac{(0.923 - 0.919)/6}{(1 - 0.922)/(570 - 16)} = 3.55$$

F critical value of 5% with dfs(6;554) is about 2.15; thus reject the null.

(f) Compute F test stastic can be computed as follows:

$$F = \frac{(1425 - (104 + 88 + ... + 211)/(70 - 16)}{(104 + 88 + ... + 211)/(570 - 70)} = 1.34$$

F critical value of 5% with dfs(54;500) is about 1.36; thus accept the null.