Student ID No. ……………………….

ECONOMICS DEPT

ASSESSED TEST

For Internal Students of
Royal Holloway

COURSE UNIT: EC5040

TITLE: Econometrics (Mid-Term Exam 1)

Date of Test 12th November 2008

Time Allowed: 1 hour

Instructions to candidates:

ANSWER BOTH QUESTIONS

WRITE ALL YOUR ANSWERS (INCLUDING ROUGH WORKING) ON THIS ANSWER BOOK

STATISTICAL TABLES ARE PROVIDED

SILENT NON-PROGRAMMABLE CALCULATORS MAY BE USED

DO NOT TURN OVER UNTIL TOLD TO BEGIN

Mid-Term Test No. 1 2008/09 – Answers

1. Given the general linear model

$$y=X\beta+ u$$

where y is an n_x1 vector of observations on the dependent variable, X is an n_xk matrix of observations on a set of explanatory variables, β is a k_x1 vector of parameters and u is an n_x1 vector of residuals

a) Derive, from first principles, an expression for the ordinary least squares (OLS) estimate of B

(8 marks)

Minimising the sum of squared residuals implies

$$Min_\beta \ \hat{u}'\hat{u} = \begin{bmatrix} \hat{u}_1 & \hat{u}_2 & .. & \hat{u}_n \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_N \end{bmatrix} = \hat{u}_1^2 + \hat{u}_2^2 + ...\hat{u}_N^2$$

$$= (y - X\hat{\beta})'(y - X\hat{\beta}) = y'y - \hat{\beta}'X'y - y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$$

Since all terms are scalars (1_x1) can add middle two terms (one is transpose of the other)

$$= y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$$

F.O.C. minimum

$$\frac{\partial \hat{u}'\hat{u}}{\partial \hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0$$

which gives k normal equations $X'X\hat{\beta} = X'y$

and the k variable OLS solution $\hat{\beta} = (X'X)^{-1}X'y$

b) Derive an expression for the variance of the OLS estimator

(6 marks)

$$Var(\hat{\beta}) = E\left[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'\right]$$

$$Var(\hat{\beta}) = E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right]$$

$= E[(X'X)^{-1}X'uu'X(X'X)^{-1}]$

$= (X'X)^{-1}X'E(uu') X(X'X)^{-1}$

$= (X'X)^{-1}X'\sigma^2 I X(X'X)^{-1}$

$\sigma^2(X'X)^{-1}$

c) Show that if the Gauss-Markov conditions are satisfied then OLS has the minimum variance of all linear unbiased estimators

(15 marks)

Consider another linear unbiased estimator $\tilde{\beta} = Cy$

If $\tilde{\beta}$ *is unbiased then* $E(\tilde{\beta}) = E(Cy) = E[CX\beta + Cu] = \beta$

Hence CX=I and $\tilde{\beta} = \beta + Cu$

So $$Var(\tilde{\beta}) = E\left[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'\right]$$

$= E[Cuu'C']$

$= \sigma^2 CC'$

Let D be the difference between the OLS and alternative estimated explanatory component ie

$$D = C-(X'X)^{-1}X'$$

So

$$Var(\tilde{\beta}) = \sigma^2\left[(D + (X'X)^{-1}X')(D + +(X'X)^{-1}X')'\right]$$

Since $CX= I = DX + (X'X)^{-1}X'X$ *the* $DX = 0$

Cross product terms vanish and

$$Var(\tilde{\beta}) = \sigma^2 DD' + \sigma^2(X'X)^{-1} = \sigma^2 DD' + Var(\hat{\beta})_{OLS}$$

ie variance of alternative estimator equals that of OLS plus a non-negative definite matrix (see problem set 0)

Hence OLS estimate has minimum variance property (BLUE – Best Linear Unbiased Estimator). Main reason for widespread use of OLS, will always provide estimators with smaller standard errors

e) Suppose that one of the independent variables is subject to a linear transformation, (multiplied by a constant λ) such that Z=XΛ where Λ is a diagonal matrix containing the transformation constant. Show the effect of this transformation on the OLS estimates of the parameters

(10 marks)

Given $y=Z\gamma + v$

OLS implies $\hat{\gamma} = (Z'Z)^{-1}Z'y$

Sub. in $Z=X\Lambda$

$$\hat{\gamma} = (\Lambda' X' X\Lambda)^{-1}\Lambda' X' y$$

Using rules on inverse of a matrix product

$$\hat{\gamma} = \Lambda^{-1}(X'X)^{-1}\Lambda'^{-1}\Lambda' X' y$$

$$\hat{\gamma} = \Lambda^{-1}(X'X)^{-1}X'y$$

$$\hat{\gamma} = \Lambda^{-1}\hat{\beta}$$

If the variable to be transformed is X_j then the transformation matrix looks like

$$\Lambda = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \lambda_j & \\ 0 & & & 1 \end{bmatrix}$$

ie a diagonal matrix with ones down the main diagonal except for the jth element which contains the constant of multiplication for the j^{th} variable

Since the inverse of a diagonal matrix is also diagonal with the reciprocal of each original element on the new main diagonal then

$$\Lambda^{-1} = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1/\lambda_j & \\ 0 & & & 1 \end{bmatrix}$$

So using the result in (3) it follows that then the corresponding regression coefficient is multiplied by 1/a and all other coefficients are unchanged.

f) Show the consequences for OLS estimation of omitting relevant variables from your model specification

(11 marks)

True: $y = X_1\beta_1 + X_2\beta_2 + e$

Estimate: $y = X_1\beta_1 + u$

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y$$
$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y + (X_1'X_1)^{-1}(X_1'X_2)\beta_2 + (X_1'X_1)^{-1}X_1'e$$
$$\hat{\beta}_1 = \hat{\beta}_1 + (X_1'X_1)^{-1}(X_1'X_2)\beta_2 + (X_1'X_1)^{-1}X_1'e$$

so

$$E(\hat{\beta}_1) = \beta_1 + (X_1'X_1)^{-1}(X_1'X_2)\beta_2 \neq \beta_1$$

OLS estimates of the coefficients on the set of X_1 variables are biased in the presence of omitted variables

and sign of bias depends on *a) the effect of the omitted variables on y, β_2,*
b) the covariance of X_1 and X_2

Not only is mean biased so is OLS estimates of parameter variances in an unknown way (If σ^2 known variance estimate is biased down. But estimate of σ^{2} 's biased up, so hard to sign direction of bias)

2. The following regression output is taken from a regression of the log of hourly pay (*lhpay*) on the years of work experience (*xper*), years of education (*yearsed*), years of job tenure, (*tenure*) a dummy variable for being female, (*female*).

Some of the regression output has been obscured.

```
reg lnhpay xper yearsed tenure female

  Source |       SS       df       MS               Number of obs =    6005
---------+------------------------------            F(  7,  6000) =
   Model |                 4                        Prob > F      =
Residual |  1600.00000  6000  .266666666            R-squared     =
---------+------------------------------            Adj R-squared =
   Total |  2400.00000  6004  .399733510            Root MSE      =  .51639
------------------------------------------------------------------------------
  lnhpay |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    xper |   .1100000   .0900000      1.222   0.212      -.0664000    .2864000
 yearsed |   .0700000   .0030894     22.861   0.000       .0645705     .076683
  tenure |  -.0100000    .030000      0.333   0.515      -.0688000    .0488000
  female |  -.1948365   .0723197     -2.694   0.007      -.3366081   -.0530649
   _cons |   .7426208   .0496538     14.956   0.000       .6452822    .8399593
------------------------------------------------------------------------------
```

a) Interpret the meaning of the coefficient on the female dummy variable

(8 marks)

this is a "semi-log" equation so the impact of being female relative to being male (net of differences in mean values of control variables) is equal to the % difference /100 in hourly pay of being female relative to being male, (since in the continuous variable case $dLnw_i/d(x) = b_i = dw_i/dw/x$) = % change in w /100 with respect to a unit change in x)

However since the coefficient is a dummy variable this is only an approximation to the proportionate difference and the true effect is closer to $\exp(\hat{\beta}_{female})-1$

So other things equal women earn exp(-.195)-1 = .177 = 17.7% less than men

b) Find the estimate of R^2 and hence test the hypothesis that the model as a whole is a good fit

(10 marks)

R^2 (the coefficient of determination) = ESS/TSS = 1 –(RSS/TSS)

From information in the regression output (highlighted in yellow)

R^2 = ESS/TSS = 1- (1600/2400) = 1 -.666 = .333

Test of goodness of fit of the model is given by $F = \frac{ess/q}{RSS/N-k} = \frac{R^2/q}{(1-R^2)/N-k} \sim F[q, N-k]$

So $F = \frac{800/4}{1600/600} = \frac{0.333/4}{0.666/6000} \sim F[4,6000]$

= 750 ~ F[4, 6000]

*So estimated F is greater than 5% critical value (F(4,∞) = 2.37) so **reject null** that model as a whole has no explanatory power*

d) The variance/covariance matrix of the OLS parameter estimates (excluding the constant) is given by

	xper	yrsed	female	tenure
xper	.00810000			
yrsed	.00024251	.00456114		
female	.00058968	.00013298	.13443127	
tenure	-.00050000	-.00670241	-.00574375	.0009000

Given this information test the hypothesis that the returns to experience (*xper*) equal the returns to job tenure (*tenure*) in the model above

(12 marks)

The test of equality of experience and tenure effects is given by a form of the F test

$$(\beta_{xper} - \beta_{tenure})^2 / Var(\hat{\beta}_{xper} - \hat{\beta}_{tenure}) \sim F(1, N\text{-}k)$$

which can check from the variance/covariance matrix of the OLS estimates since the square root of the i^{th} element on the main diagonal should equal the standard error on the ith variable in the regression and the off diagonal terms are the covariances of the parameter estimates. The relevant covariance is highlighted (in green)

Since $Var(\hat{\beta}_{xper} - \hat{\beta}_{tenure}) = Var(\hat{\beta}_{xper}) + Var(\hat{\beta}_{tenure}) - 2Cov(\hat{\beta}_{xper}\ \hat{\beta}_{tenure})$

$Var(\hat{\beta}_{xper} - \hat{\beta}_{tenure}) = .0081 + .0009 - 2(-.0005) = .01$

then $F = \dfrac{(.11 - -.01)^2}{.01} = 1.44 \qquad \sim F[1, 6000]$

Since 95% critical value = 3.84 then $\hat{F} < F_{critical}$

So can **not** reject null that coefficients are equal (the relatively large standard errors ensure that the confidence intervals overlap)

d) Consider a simple model of 204 observations split equally into two sub-samples such that

$y_i = a_1 + b_1 X_i + u_i \qquad i=1..N_1 \qquad$ in sub-sample 1

and

$y_i = a_2 + b_2 X_i + u_i \qquad i=N_1+1..N \qquad$ in sub-sample 2

Suppose that $RSS_1 = 8$ and $RSS_2 = 2$ and that the RSS from the pooled regression is 12. Test the hypothesis of no structural change across the two sub-samples at the 5% level.

(10 marks)

The unrestricted form of the model (intercepts and the slopes vary in two periods) in (partitioned) matrix form is given by

$$y = \begin{bmatrix} y_1 \\ .. \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & : & 0 \\ .. & .. & .. \\ 0 & : & X_2 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ .. \\ u_2 \end{bmatrix} = X\beta + u \qquad (1)$$

where X_1 is an N_1 by 2 matrix of observations from the 1^{st} sub-sample and X_2 is an N_2 by 2 matrix of observations from the 1^{st} sub-sample with $N = N_1 + N_2$

ie stacking the data from the second period below that of the observations from the 1st period in a way that allows the coefficients to differ between the periods

Compare this with estimates from the restricted (pooled) model based on

$$y = \begin{bmatrix} y_1 \\ .. \\ y_2 \end{bmatrix} = \begin{bmatrix} i\, X_1 \\ i\, X_2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} u_1 \\ .. \\ u_2 \end{bmatrix} = X\beta + u$$

To test formally use

$$F = \frac{(RSS_{restricted} - RSS_{unrestricted})/q}{RSS_{unrestricted}/N-k} \sim F[q, N-k]$$

which in this case becomes the Chow test

$$F = \frac{(RSS_{restricted} - RSS_1 + RSS_2)/q}{RSS_1 + RSS_2 / N - 2k} \sim F[q, N-2k]$$

(remember that there are 4 parameters in the unrestricted model so k=4 and q= 2 restrictions

hence $\hat{F} = \dfrac{(12-(8+2))/2}{(8+2)/204-2*2} = 20$

From Tables the 5% critical value given the degrees of freedom $F^{.05}[2,\infty] = 3.0$

$\hat{F} > F_{critical}$ *so reject null (of no structural change)*

d) Outline the form of a technique that could be used to test for functional form error (10 marks)

Either

Ramsey RESET Test
- if model is good fit then addition of extra variables should not be statistically significant

rather than add higher order terms of original variables a more parsimonious alternative is to use fact that

$$\hat{y} = X\hat{\beta}$$

so predicted values are linear function of all the X variables (weighted by their estimated coefficients)

and hence $(\hat{y})^j = (X\hat{\beta})^j$

are linear functions of higher powers of all the X variables

$$y = X\beta + \delta_2 \hat{y}^2 + \delta_3 \hat{y}^3 + \ldots + \delta_j \hat{y}^j + u$$

and test null H0: $\delta_2 = \delta_3 = \ldots \delta_j = 0$

If estimated F value greater than critical value reject null that functional form is acceptable.

OR

LM Test of Omitted Variables

1. Run restricted regression (no higher order terms)

2. save residuals

3. Regress residuals on unrestricted model (containing higher order values of X (or the $\hat{y}_j$) - the auxiliary regression

Can show

$$NR^2_{aux} \overset{a}{\sim} \chi^2_{(No. of restrictions)}$$

If estimated Chi-squared value greater than critical value reject null that functional form is acceptable.