STUDENT ID: _____

HEMIS No._____

For Internal Students of
Royal Holloway

DO NOT TURN OVER UNTIL TOLD TO BEGIN

EC2203: ECONOMETRICS
EC4203: ECONOMETRICS

Mid-Term Examination No. 3

Time Allowed: 55 minutes

Answer Both questions

STATISTICAL TABLES ARE PROVIDED

Silent non-programmable calculators may be used

PRINT YOUR STUDENT NUMBER ON THE FRONT OF THIS TEST PAPER WHERE INDICATED

WRITE ALL YOUR ANSWERS (INCLUDING ROUGH WORKING) ON THIS TEST PAPER. THERE ARE EXTRA BLANK SHEETS TOWARD THE BACK OF THE PAPER

1. The following output is taken from an OLS regression of the natural log of food expenditure (*lfood)* on the level of household income, measured in £1000 a week, (*hincome)*

Some of the regression output has been hidden deliberately.

```
. reg lfood  hincome

      Source |       SS       df       MS              Number of obs =
-------------+------------------------------           F(  1,   100) =
       Model |                 1                        Prob > F      =  0.0065
    Residual |    9.000000   100                        R-squared     =
-------------+------------------------------           Adj R-squared =
       Total |  12.0000000   101   .11881188           Root MSE      =  .34461

------------------------------------------------------------------------------
       lfood |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     hincome |    .200000   .1000000
       _cons |   3.738375   .0591427    63.21   0.000     3.621008    3.855741
------------------------------------------------------------------------------
```

a) Find the sample size

(5 marks)

- Since we know in a regression output that the regression degrees of freedom are calculated as N-k

where N is the sample size and k is the number of right hand side parameters, then can see from the above that

100 = N- k = N-2 so N=102

b) Calculate the R^2 value

(5 marks)

Need R^2 = 1-RSS/TSS = (1- 9)/12 = 0.25 *or R^2 =ESS/TSS*
where ESS = 12-9=3
and 3/12 = 0.25

c) Test the hypothesis that the variable *income* has some explanatory power (use the 95% significance level for your test)

(5 marks)

Use $\hat{t} = \frac{\hat{\beta} - \beta^0}{s.e.(\hat{\beta})}$ *=(0.2-0)/0.1= 2*

From tables critical value at the 95% level given N-k =102-2 =100 degrees of freedom (2-tailed test) is 1.98 (Not 1.96)

Hence absolute value of estimated t > $t_{critical}$ so reject null hypothesis ***that variable has no explanatory power***

Note that in this case with only 1 explanatory variable – the t test and the F test if goodness of fit will be equivalent (The F value will be the square of the t value. SO could use F test as an alternative answer)

d) Find the 99% confidence interval in which the true value of the coefficient on *income* will lie.

(5 marks)

use $\Pr\left[\hat{\beta}_1 - t_{N-k}^{\alpha/2} * s.e.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{N-k}^{\alpha/2} * s.e.(\hat{\beta}_1)\right] = 0.99$

so can be 99% confident that true value lies in range

*0.2– (2.62*0.1) <=β <= 0.2 + (2.62*0.1)*
-0.062 <= β <= 0.462

(Note that 99% confidence interval – unlike the 95% confidence interval does not exclude the possibility that the true effect of the variable could be zero)

e) Interpret the estimated effect of household income on food expenditure

(5 marks)

This is a log-lin model so the coefficients are semi-elasticities

so estimated effect is $\frac{dLog(Food)}{dincome} = \beta_{income}$

which is the definition of a semi-log elasticity and the effect can be interpreted as the percentage change in food expenditure with respect to a unit change in income Since

$$dLn(Food) = \frac{dFood}{Food} = \%\Delta food \div 100$$

Then given $\%\Delta Food \div 100 = (\beta_{income} * dIncome)$

and so $\%\Delta Food = (\beta_{income} * dIncome) * 100$

so a £1000 rise in income ***increases*** *food expenditure by 0.2*1*100 %= 20%*
and a £10,000 rise in income ***increases*** *food expenditure by 0.2*10*100 %= 200%*

(remember the initial units of measurement was in £1000)

f) Calculate the approximate p value for the estimated coefficient on *income*

(5 marks)

Since p value is the significance level which makes the estimated t value equal to the critical value then need to look at t tables to find which size test puts an estimated t value of 3 at the border with 100 degrees of freedom. Can see lies between .05 and .02 (on the 2 sided test), and nearest value in Table is .05 (though is probably closer to something like .048)

g) Calculate the OLS estimate of the residual variance of the model

(5 marks)

$s^2 = RSS/N\text{-}k = 9/100 = 0.9$

h) Find the variance of income

(7 marks)

Since this is a two variable model we know that the standard error of income has been estimated by the formula

$$s.e.(\hat{\beta}_1) = \sqrt{\frac{s^2}{N * Var(X)}}$$

substituting in the values from the information acquired so far then

$$0.1 = \sqrt{\frac{0.9}{102 * Var(Income)}}$$

so $(0.1)^2 * (102Var(Income)) = 0.9$

Var(Income) = 0.9/(102*.0.01) = 0.88

i) What happens to the estimated OLS coefficient on income if the income variable is rescaled and is now measured in £ rather than in £1000?

(8 marks)

Since the OLS estimate of the coefficient on income is given by

$$\hat{\beta}_{income} = \frac{Cov(LnFood, Income)}{Var(Income)}$$

then the rescaling means that the income variable is multiplied by 1000 and so
(1 under the old measure is 1000 under the new, 5 is now 5000 etc)

$$\hat{\beta}_{income} = \frac{Cov(LnFood, Income * 1000)}{Var(Income * 1000)}$$

which using rules on covariances implies that

$$\hat{\beta}_{income} = \frac{Cov(LnFood, Income)/*1000}{Var(Income) * 1000000} = \frac{Cov(LnFood, Income)}{Var(Income)} * \frac{1}{1000} = \hat{\beta}_{income}/1000$$

so the OLS coefficient estimate is rescaled by the reciprocal of the coefficient of rescaling

2.

a) What are the 5 main Gauss-Markov assumptions concerning the behaviour of the residuals that underlie the OLS estimation technique?

(15 marks)

i) $E(u_i) = 0$ average value of true (unobserved) residuals is zero – nothing systematic
ii) $Var(u_i) = E(u_i^2/X_i) = constant$
ie variance of residuals is constant for any value of X (homoskedasticity)
iii) $Cov(u_i\ u_j) = 0$ for all I not equal to j (zero autocorrelation)
value of one residual gives no information about value of another
iv) No covariance between X and true residual $Cov(X, u) = 0$
v) The residuals are normally distributed $\sim N(0, \sigma_u^2)$

b) What does the Gauss-Markov theorem say about OLS estimation?

(5 marks)

*- that **if** the 5 assumptions above hold the OLS will produce unbiased estimates that have the smallest variance of any other (linear) unbiased estimation techniques. Normality in the residuals also ensures that the OLS estimates will be normally distributed and that the linear hypothesis tests will be based on t and F distributions*

c) What do you understand by the term "forecast error"?

(5 marks)

Given a set of in-sample OLS estimates of the intercept and the slope then for ANY value of X (in or out of sample) it is possible to produce a prediction.

$$\hat{u}_0 = \hat{y}_o - y_o$$
$$= \hat{\beta}_0 + \hat{\beta}_1 X_o - y_o$$

(o is "out of sample")

the closer is the estimate $\hat{y}$ to its actual value, the smaller will be the forecast error

d) Outline the form of the Chow Forecast Test for Out-of-Sample Parameter stability

(7 marks)

If a model forecasts well out of sample then we would expect all the out-of-sample residuals to be close to zero. It can be shown that the joint test of all the out-of-sample-residuals being close to zero is given by:

$$F = \frac{RSS_{in+out} - RSS_{in} / N_o}{RSS_{in} / N - k} \sim F[N_o, N - k]$$

where *N_o is the number of out-of-sample observations*
N is the number of in-sample observations
k is the number of RHS coefficients

Intuitively if the model fits well the RSS from the combined regression should be close to that from the in-sample regression. A "large" difference suggest the RSS are different and so model does not forecast well)

Given a null hypothesis that the model is stable out of sample (predicts well) then if

$$\hat{F} > F_{critical}{}^{\alpha}[N_o, N-k]$$

***reject** null of model stability out-of-sample*

e) What do you understand by the term multicolinearity?

(6 marks)

In the 3 variable model can show that

$$Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$$

r^2_{x1x2} is the square of the correlation coefficient between X_1 & X_2

(compared with $Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)}$ in the 2 variable model)

So an increased correlation between X_1 & X_2 will make the OLS estimates of the effects of the X variables less precise (can't distinguish between the contribution of the individual variables if correlation is high)

The high correlation is called ***multicolinearity***

f) What is the consequence of multicolinearity for OLS estimation?

(5 marks)

the symptoms are that

1) while OLS estimates remain unbiased
2) the standard errors are much larger than would be in the absence of multicolinearity

and since $\hat{t} = \frac{\hat{\beta}_1 - \beta_1^{\ 0}}{s.e.(\hat{\beta}_1)}$

the estimated t values will be smaller than otherwise.

You may therefore conclude that variables are statistically insignificant (from zero) when not (ie Type II error)

g) How would you detect the presence of multicolinearity?

(7 marks)

Detection:

1) Low t values and high R^2
2) The estimates may be sensitive to addition or subtraction of a small number of observations

Look at the simple correlation coefficients between any 2 variables. A correlation coefficient >0.8 usually says there are problems. Or if the correlation between any two right hand side variables is greater than the correlation between that of each with the dependent variable. In case with many right hand side variables run an auxiliary regression of *any one* of the right hand side variables on *all* the other X variables

$$X1 = \delta_0 + \delta_2 X_2 + \delta_3 X_3 + \ldots \delta_k X_k + u$$

and look at the R^2 from this regression. An $R^2 > 0.8$ suggests problems