STUDENT ID: ________________________

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: 1 hour

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 OLS regressions of three different models which try to establish the effect of household income (measured in £ a week) on food expenditure (measured in £ a week). The first model regresses the level of food expenditure, (food), on the level of household income, (*hhincome)*. The second model regresses the natural log of food expenditure, (*lfood),* on the natural log of household income, (*lhhincome)* and the third model regresses the natural log of food expenditure, (*lfood),* on the level of household income, (*hhincome).*

Some of the regression output has been hidden deliberately.

```
1.  reg food hhincome

      Source |       SS       df       MS              Number of obs =
-------------+------------------------------           F(  1,    30) =    4.00
       Model |  14946.8281     1  14946.8281           Prob > F      =  0.0000
    Residual |  49088.0637    30  247.919514           R-squared     =  0.2334
-------------+------------------------------           Adj R-squared =  0.2295
       Total |  64034.8918    31  321.783376           Root MSE      =  15.745
------------------------------------------------------------------------------
        food |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    hhincome |   .0180000   .0090000
       _cons |   31.47472   1.643727    19.15   0.000     28.23326    34.71617
------------------------------------------------------------------------------

2. reg lfood lhhincome

      Source |       SS       df       MS              Number of obs =
-------------+------------------------------           F(  1,    30) =  113.73
       Model |  16.7099173     1  16.7099173           Prob > F      =  0.0000
    Residual |  29.0921201    30  0.96973734           R-squared     =  0.3648
-------------+------------------------------           Adj R-squared =  0.3616
       Total |  45.0000000    31  1.45161290           Root MSE      =  .98475
------------------------------------------------------------------------------
       lfood |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   lhhincome |   .3725856   .0349377    10.66   0.000     .3036879    .4414833
       _cons |   1.368225     .21143     6.47   0.000     .9512815    1.785169
------------------------------------------------------------------------------

3. reg lfood hhincome

      Source |       SS       df       MS              Number of obs =
-------------+------------------------------           F(  1,    30) =
       Model |                 1                       Prob > F      =  0.0000
    Residual |  36.0000000    30                       R-squared     =
-------------+------------------------------           Adj R-squared =
       Total |  45.0000000    31  1.45161290           Root MSE      =  1.0954

------------------------------------------------------------------------------
       lfood |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    hhincome |    .000437
       _cons |   3.365214   .0445782    75.49   0.000     3.277305    3.453123
------------------------------------------------------------------------------
```

a) Interpret the effect of the estimated coefficient on the explanatory variable in **each** of the three models

(9 marks)

continued over

Model 1 is a linear model so the coefficients can be interpreted as

1st equation is in levels so estimated effect is $\frac{dFood}{dHHIncome} = \beta_{hhincome}$

and the change in food expenditure is given by $dFood = \beta_{hhincome} * dHHincome$

*so a £1 rise in weekly household income, increases weekly food expenditures by 0.017*1 1.7 pence (and a £10 increase raises food expenditure by 17 pence)*

Model 2 equation is log-linear so estimated effect is $\frac{dLog(Food)}{d\log(HHincome)} = \beta_{\log(HHincome)}$

which is the definition of an (income) elasticity and the effect can be interpreted as the percentage change in food expenditure wrt a x% change in household income given by $d\log(Food) = \beta_{\log(hhincome)} * d\log(Hhincome)$

*so a 1% rise in income, increases food expenditure by 0.37*1 = 0.37%*

(Note in order to use this model both y and X variables should always be positive – can't take (natural) log of a negative number)

Note also that the implied slope of this predicted line of Y against X changes at very value of X, but the elasticity is constant at every value of X

Model 3 is log-lin (semi log) so estimated effect is $\frac{dLog(Food)}{dhhincome} = \beta_{hhincome}$

which is the definition of a semi-log elasticity and the effect can be interpreted as the percentage change in food expenditure wrt a unit change in income given by $d\log(Food) = (\beta_{hhincome} * dHhincome)*100$*)*

*In this case a unit is £1, so a £1 rise in income, increases food expenditure by 0.0004*1 = 0.04%*

b) Find the sample size in model 1

(5 marks)

- Since we know in a regression output that the regression degrees of freedom are calculated as N-k and the F test of goodness of fit is also distributed Ff[q, N-k] where N is the sample size and k is the number of right hand side parameters and q Is the number of restrictions, then can see from the above output on the F test that

30 = N- k and since in this model k=2 then 30= N-2 so N=32

c) Calculate the R^2 value in model 3

(5 marks)

Need R^2 *= 1-(RSS/TSS) = 1-(36/45) = 1-0.8 = 0.2*

Equivalently R^2 *=ESS/TSS where ESS = 45-36=-9*

and 9/45 = 0.2

d) Test the hypothesis that the variable *hhincome* has no explanatory power in model 1 (use the 95% significance level for your test)

(5 marks)

Use $\hat{t} = \dfrac{\hat{\beta} - \beta^0}{s.e.(\hat{\beta})}$ *=(0.018-0)/0.009= 2*

From tables nearest critical value at the 95% level given N-k = 32-2 = 30 degrees of freedom (2-tailed test) ***is 2.04*** *(**not** 1.96)*

*Hence **absolute** value of estimated t < $t_{critical}$ (2<2.04) so **accept** 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 of 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)

e) Find the 95% confidence interval for the true value of the coefficient on *hhincome* in model 1

(5 marks)

Since α=0.05 (5%) and this is a 2-tailed test then the confidence interval is given by

$$\Pr\left[\hat{\beta}_1 - t_{N-k}^{.05/2} * s.e.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{N-k}^{.05/2} * s.e.(\hat{\beta}_1)\right] = 0.95$$

so can be 95% confident that true value lies in range

*0.018– (2.04*0.009) <=β <= 0.018 + (2.04*0.009)*
-0.00036 <= β <= 0.03636

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

f) Calculate the approximate p value for the estimated coefficient on *income in model 1*

(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 2 at the border with 30 degrees of freedom. Can see lies between .10 and .05 (on the 2 sided test), and nearest value in Table is .05 (though is probably closer to something like .052)

g) Calculate the OLS estimate of the residual variance in model 3

(5 marks)

s^2 = RSS/N-k = 36/30 = 1.2

h) If the variance of the variable *hhincome* is 100, find the OLS estimate of standard error on the coefficient of *hhincome* in model 3

(6 marks)

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

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

substituting in the values from the information acquired so far then

$$0.1 = \sqrt{\frac{1.2}{32*100}}$$

$$s.e.(\hat{\beta}_{hhincome}) = 0.019$$

i) What happens to the estimated OLS coefficient on income in model 1 if the income variable is rescaled and is now measured in pence rather than in £ ?

(9 marks)

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

$$\hat{\beta}_{hhincome} = \frac{Cov(Food, HHincome)}{Var(HHincome)}$$

then the rescaling means that the income variable is multiplied by 100 and so (1 under the old measure is 100 under the new, 5 is now 500 etc)

$$\hat{\beta}_{hhincome} = \frac{Cov(Food, Income*100)}{Var(Income*100)}$$

which using rules on covariances implies that

$$\hat{\beta}_{hhincome} = \frac{Cov(Food, Income)*100}{Var(Income)*10000} = \frac{Cov(LnFood, Income)}{Var(Income)*100} = \frac{\hat{\beta}_{hhincome}}{100}$$

*so the original OLS coefficient estimate is **divided by 100** if the units of measurement of **the right hand side** variable is changed by the same amount*

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) The following is regression output from a model that seeks to explain the number of prisoners (*inmates*) by the unemployment rate (*urate).* The model is estimated on the first 100 observations with 10 observations held back.

```
i) reg  inmates urate if _n<=100
      Source |       SS       df       MS              Number of obs =     122
-------------+------------------------------           F(  1,   120) =  285.44
       Model |   6201.9895     1   6201.9895           Prob > F      =  0.0000
    Residual |   3000.0000   120   25.000000           R-squared     =  0.6739
-------------+------------------------------           Adj R-squared =  0.6763
       Total |  9201.98950   121   76.049000           Root MSE      =  5.0000
------------------------------------------------------------------------------
     inmates |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       urate |  -3.239703   .1917564   -16.89   0.000    -3.619098   -2.860309
       _cons |   75.64658   1.353774    55.88   0.000     72.96811    78.32506
------------------------------------------------------------------------------
```

```
ii) reg  inmates urate

      Source |       SS       df       MS              Number of obs =     132
-------------+------------------------------           F(  1,   120) =  932.64
       Model |   13015.548     1   13015.548           Prob > F      =  0.0000
    Residual |    5000.000   130   38.461538           R-squared     =  0.7224
-------------+------------------------------           Adj R-squared =  0.7211
       Total |   18015.548   131  137.523270           Root MSE      =  4.9677
------------------------------------------------------------------------------
     inmates |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       urate |  -4.274686   .1399738   -30.54   0.000    -4.550683   -3.998688
       _cons |   83.78884    .830109   100.94   0.000     82.15205    85.42563
------------------------------------------------------------------------------
```

Outline the intuition that underlies the Chow forecast test of parameter stability

(4 marks)

If a model forecasts well out of sample then we would expect all the out-of-sample residuals to be close to zero. 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*

Do the formal test given the information in the regression output above

(6 marks)

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

is given by

$$F = \frac{(5000-3000)/10}{3000/120} = 8$$

The 95%critical value at the relevant degrees of freedom is the same as before

. F(10,120)=1.91

so $\hat{F} > F_{critical}{}^{\alpha}[N_O, N-k]$ *so **reject null** that model predicts well out of sample*

e) The following output describes the distribution of OLS regression residuals

```
. su reshat1 if _n<201, det
                          Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -6.323713      -6.448029
 5%    -5.227664      -6.414042
10%    -4.487434      -6.233384       Obs                 180
25%    -3.299453       -6.18045       Sum of Wgt.         180
```

50%	-.9010892		Mean	-1.44e-10
		Largest	Std. Dev.	4.585193
75%	1.951943	16.25964		
90%	5.31512	17.23212	Variance	21.02399
95%	9.095832	17.61862	Skewness	2.00000
99%	17.42537	20.11699	Kurtosis	5.00000

Describe the 2 attributes used to determine whether a variable is likely to follow a normal distribution

(6 marks)

Residuals show signs of right skewness (residuals bunched to left, the median is less than the mean of zero – so not symmetric) Skewness value is above zero – which again indicates right skewness

and kurtosis (leptokurtic – since peak of distribution higher than expected for a normal distribution where kurtosis is equal to 3)

Why might we worry if the OLS residuals are not normally distributed?

(4 marks)

The assumption of normality is needed to derive the t test, confidence intervals and F tests need for hypothesis testing. If residuals not normal then t values may not follow a t distribution

Hence do a formal test of normality in these residuals.

(7 marks)

To test more formally construct Jarque-Bera test

$$JB = N * \left[\frac{Skewness^2}{6} + \frac{(Kurtosis-3)^2}{24} \right]$$

$jb = (180/6)*((2^2)+(((5-3)^2)/4)) = 150$

*The statistic has a Chi^2 distribution with **2** degrees of freedom, (one for skewness one for kurtosis).*

From tables critical value at 5% level for 2 degrees of freedom is 5.99

*So JB>$\chi^2_{critical}$, so **reject** null that residuals are normally distributed.*

Suggests should try another functional form to try and make residuals normal, otherwise t stats may be invalid.