UNIVERSITY COLLEGE LONDON

EXAMINATION FOR INTERNAL STUDENTS

MODULE CODE	:	**ECON2007**
ASSESSMENT PATTERN	:	**ECON2007A**
MODULE NAME	:	**Quantitative Economics and Econometrics**
DATE	:	**06 May 2016**
TIME	:	**10:00 am**
TIME ALLOWED	:	**3 hours**

This paper is suitable for candidates who attended classes for this module in the following academic year(s):

2013/14, 2014/15 and 2015/16

SUMMER TERM 2016

ECON2007: QUANTITATIVE ECONOMICS AND ECONOMETRICS

TIME ALLOWANCE: 3 hours

Answer all 3 questions from Part A and Answer 1 question from Part B.

Questions in Part A carry 60 *per cent of the total mark and questions in Part B carry* 40 *per cent of the total mark each.*

In cases where a student answers more questions than requested by the examination rubric, the policy of the Economics Department is that the student's first set of answers up to the required number will be the ones that count (not the best answers). All remaining answers will be ignored.

PART A

Answer all questions from this section.

A.1 True or False, and explain your answer

(a) If $y \sim N(\mu, 1)$ (where $\bar{y}$ is the sample mean of y), $\frac{1}{\bar{y}}$ is an unbiased and consistent estimator of $\frac{1}{\mu}$.

(b) If x and y are independent of each other, then $Cov(x, y) = 0$.

A.2 Table 1

```
. reg lwage educ

      Source |       SS       df       MS              Number of obs =    1084
-------------+------------------------------           F(  1,  1082) =  145.88
       Model |  37.9092742     1  37.9092742           Prob > F      =  0.0000
    Residual |  281.181893  1082  .25987236            R-squared     =  0.1188
-------------+------------------------------           Adj R-squared =  0.1180
       Total |  319.091167  1083  .29463635            Root MSE      =  .50978

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0691514   .0057254    VVVV   XXXX       YYYY         ZZZZ
       _cons |   .9839614   .0747576   13.16   0.000     .8372751    1.130648
------------------------------------------------------------------------------
```

Using data from the US Current Population Survey, an economist estimates the regression in table 1 above, where lwage is the natural log of the wage, and educ is years of education.

(a) Calculate VVVV, YYYY, and ZZZZ in the table above, and make a good guess for the likely value of XXXX. Interpret the p-value associated with the educ.

(b) Next the same economist estimates the regressions in table 2 below.
Table 2

```
. reg lwage

      Source |       SS       df       MS              Number of obs =    1084
-------------+------------------------------           F(  0,  1083) =    0.00
       Model |           0     0           .           Prob > F      =       .
    Residual |  319.091167  1083  .29463635            R-squared     =  0.0000
-------------+------------------------------           Adj R-squared =  0.0000
       Total |  319.091167  1083  .29463635            Root MSE      =   .5428

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   1.867301   .0164865   113.26   0.000     1.834952     1.89965
------------------------------------------------------------------------------

. predict lwage_pred
(option xb assumed; fitted values)

. gen lwage_dev=lwage-lwage_pred

. reg educ

      Source |       SS       df       MS              Number of obs =    1084
-------------+------------------------------           F(  0,  1083) =    0.00
       Model |           0     0           .           Prob > F      =       .
    Residual |  7927.62638  1083  7.3200613            R-squared     =  0.0000
-------------+------------------------------           Adj R-squared =  0.0000
       Total |  7927.62638  1083  7.3200613            Root MSE      =  2.7056

------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   12.77399   .0821756   155.45   0.000     12.61274    12.93523
------------------------------------------------------------------------------

. predict educ_pred
(option xb assumed; fitted values)

. gen educ_dev=educ-educ_pred
```

The economist then estimates

```
. reg lwage_dev educ_dev
```

Explain what would the estimated coefficients be, both for the intercept and the coefficient on educ_dev. Give numbers for the estimated coefficients, and use derivations for why we should get these numbers. **Hint:** The regression "reg lwage" is the following regression $lwage_i = \beta_0 + u_i$. Thus $\hat{\beta}_0$ is the sample average of *lwage*.

(c) The economist then estimates

```
. reg lwage educ_dev
```

Explain what would the estimated coefficients be, both for the intercept and the coefficient on educ_dev. Give numbers for the estimated coefficients, and derive your results.

(d) The economist suspects that the effect of education varies with gender, so she adds in female (which is a 0-1 indicator =1 if female) and female $\times$ educ. So she obtains the estimates in table 3 below:

Table 3

```
. gen female_educ=female*educ

. reg lwage educ female female_educ

      Source |       SS       df       MS              Number of obs =    1084
-------------+------------------------------           F(  3,  1080) =   79.89
       Model |  XXXXXXXXXX     3  XXXXXXXXXX           Prob > F      =  0.0000
    Residual |  XXXXXXXXXX  1080  XXXXXXXXXX           R-squared     =  0.1816
-------------+------------------------------           Adj R-squared =  0.1793
       Total |  319.091167  1083   .29463635           Root MSE      =  .49173

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0654327   .0066599     9.82   0.000     .0523648    .0785006
      female |  -.5090672   .1562024    -3.26   0.001    -.8155617   -.2025726
 female_educ |   .0184717   .0119386     1.55   0.122    -.0049538    .0418973
       _cons |   1.144336   .0866862    13.20   0.000     .9742431    1.314428
------------------------------------------------------------------------------
```

Compare tables 1 and 3. Can we reject the hypothesis that the coefficients on female and female_educ are statistically different from 0 at the 5% level? To solve this it might be helpful to know the following 5% critical values for the F-distribution: $F_{1,1080} = 3.89, F_{2,1080} = 3.04, F_{3,1080} = 3.65, F_{4,1080} = 2.42$.

(e) Next, she estimates

```
. gen male_educ=male*educ

. reg lwage educ male male_educ
```

where male is a 0-1 indicator and is equal to if male. What will the estimates be?

(f) Suppose that age of the individual was added to the regression of lwage on educ, and that older people tend to have lower levels of education attainment. Would you anticipate the coefficient on educ to rise or fall? Why? Explain in words.

A.3 Keynes suggested a consumption function of the form

$$C_t = \beta_0 + \beta_1 Y_t + u_t$$

where C_t and Y_t are consumption and income at time t.

(a) Keynes hypothesized that the marginal and average propensity to consume were both positive, and that the average propensity to consume was falling in income. What restrictions does this impose on β_0, β_1?

(b) Suppose that he estimated $\hat{\beta}_0 = 1, \hat{\beta}_1 = .9, \widehat{Var(u)} = 1$ (i.e., the estimated variance of the residuals is 1) , $\widehat{Var(Y)} = 10$ (i.e., the estimated variance of Y is 10). Report $\hat{C}$ (predicted consumption) as a function of Y, and $Var(\hat{C})$.

(c) What is the R^2 of the above regression?

(d) Suppose instead he decided to estimate savings, rather than consumption, as a function of income, where savings is defined as $S_t = Y_t - C_t$. He then estimates

$$S_t = \delta_0 + \delta_1 Y_t + e_t$$

What is $\hat{\delta}_0, \hat{\delta}_1$? What is the R^2 of this regression?

PART B

Answer ONE question from this section.

B.1 In "Competition and Innovation: An Inverted-U Relationship" (*Quarterly Journal of Economics*, Vol.120, No.2, 2005), Philippe Aghion, Nick Bloom, Richard Blundell, Rachel Griffith and Peter Howitt study the relationship between competition and innovation. This question is based on the specifications in the article, which use measures for both variables and investigates their empirical association. The innovation measure used in the study is (essentially) the number of patents issued in an industry during a given year over the period 1973 to 1994. To register the degree of competition in an industry, the authors use one (1) minus a measure called the Lerner index. An industry in perfect competition would have the index equal one. Lower values for the index register deviations from perfect competition.

(a) Assuming that a cross-section of industries is available, the number of patents p_j in industry j can be modelled as a random variable with Poisson distribution. Given the competition index c_j in industry j, the authors estimate a Poisson regression where the number of patents in industry j depends on the competition variable c_j in that industry. A (simpler) version of their Poisson regression model specifies the expected number of patents given the competition variable as

$$\mathbb{E}(p_j|c_j) = \exp(\beta_0 + \beta_1 c_j + \beta_2 c_j^2).$$

Using their data, the point estimates for the parameters above are:

$$\hat{\beta}_0 = -75.01 \quad \hat{\beta}_1 = 165.12 \quad \hat{\beta}_2 = -88.55.$$

The sample average of c_j is 0.95. Compute the Partial Effect at the Average (PEA). (Show your calculations.) The OLS estimates for the coefficients in the linear regression

$$p_j = \alpha_0 + \alpha_1 c_j + \alpha_2 c_j^2 + u_j,$$

on the other hand, are

$$\hat{\alpha}_0 = -474.20 \quad \hat{\alpha}_1 = 1032.86 \quad \hat{\alpha}_2 = -554.08.$$

and give a PEA of -19.91. How do the PEA from the Poisson regression and the PEA from the OLS compare? (Explain any differences.)

(b) How do you interpret the estimated PEA in the Poisson regression above? The standard error for the estimate of this PEA is 7.43. Describe how you would obtain a confidence interval for the PEA at a particular confidence level (e.g., 95%).

(c) Suppose now that you only observe whether a particular industry j issues any patents or none in a given year (i.e., $p_j = 0$ or $p_j > 0$). Remember that the Poisson regression model stipulates that

$$\mathbb{P}(p_j = k|c_j) = \frac{\exp(\beta_0 + \beta_1 c_j + \beta_2 c_j^2)^k \exp[-\exp(\beta_0 + \beta_1 c_j + \beta_2 c_j^2)]}{k!},$$

for $k = 0, 1, 2, \ldots$ A classmate suggests estimating this model by Maximum Likelihood. Write down the (log-)likelihood function for this estimation problem taking into account the fact that one only observes $p_j = 0$ or $p_j > 0$.

(d) The authors point out that high levels of innovation may reduce competition and this would make c_j endogenous. To address this problem, they argue they employ an instrumental variable that reflects policy changes affecting competition. Those policies are the Thatcher era privatisations, the EU Single Market Programme and the Monopoly and Merger Commission investigations that resulted in structural changes in certain industries. Let z_j be a single variable encoding these policy instruments and suppose that it is related to c_j according to:

$$c_j = \pi_0 + \pi_1 z_j + v_j.$$

Assume that the regression of interest is linear in c_j so that:

$$p_j = \beta_0 + \beta_1 c_j + u_j.$$

Describe how you would implement the TSLS estimator in this context. How would you argue for the validity of this instrument?

(e) Suppose you are interested in modelling the evolution of the number of patents in a given industry in year t, p_t, using the following autoregressive model:

$$\ln p_t = \rho_0 + \rho_1 \ln p_{t-1} + \eta_t, \qquad |\rho_1| < 1.$$

If you estimate the model by OLS, is the estimator unbiased? Why? Is it consistent? Why?

(f) How would you test whether there is serial correlation in η_t?

B.2 This question is based on "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate", K.C. Chan, G. Andrew Karolyi, Francis A. Longstaff and Anthony B. Sandres (*The Journal of Finance*, 1992). In this article, the authors are interested in a regression model for the short-term interest rate r_t given by:

$$r_t - r_{t-1} = \beta_0 + \beta_1 r_{t-1} + \epsilon_t.$$

Assume that the errors are not serially correlated.

(a) Provide sufficient conditions (on β_1) for r_t to be stationary and ergodic. In this case, will the OLS estimator of a linear regression of r_t on r_{t-1} be unbiased? Explain.

(b) A few well-known models in finance hypothesize that $\beta_1 = 0$. Explain how would you test this hypothesis.

Suppose you are interested in the following relation between long-term interest rates R_t and short-term interest rates r_t:

$$R_t - r_t = \alpha_0 + \alpha_1 r_t + \eta_t. \tag{1}$$

(c) Assume that $r_t = \beta_0 + \epsilon_t$ (i.e., $1 + \beta_1 = 0$). If the data is generated according to the model above (i.e., equation (1)), under what conditions is the OLS estimator for α_0 and α_1 consistent?

(d) Assume that $r_t = \beta_0 + \epsilon_t$ (i.e., $1 + \beta_1 = 0$). To obtain an estimate for α_1, one can first obtain the residuals from a regression of r_t on a constant, call it $\tilde{r}_t$, and the residuals from a linear regression of $R_t - r_t$ on a constant, call it $\widetilde{R_t - r_t}$, and obtain OLS estimates for a linear regression of $\widetilde{R_t - r_t}$ on $\tilde{r}_t$ (and a constant). How does the slope coefficient estimator obtained compare with the OLS estimator applied to equation (1)? Explain.

(e) Assume that $r_t = \beta_0 + \epsilon_t$ (i.e., $1 + \beta_1 = 0$). How would you test whether η_t is serially correlated? Explain.

(f) Assume now that $\beta_1 = 0$. If η_t is stationary, what can you say about the OLS estimator for α_0 and α_1 applied to equation (1)? What if it is not stationary? Explain.