AFRICAN ECONOMIC RESEARCH CONSORTIUM

Collaborative MA Programme in Economics for Anglophone Africa (Except Nigeria)

JOINT FACILITY FOR ELECTIVES (JFE) 2010

JUNE – OCTOBER

ECONOMETRICS THEORY AND PRACTICE II

Second Semester: Final Examination

Duration: 3 Hours **Date: Monday, September 27, 2010**

INSTRUCTIONS:

1. Answer **Question 1** and **ANY TWO (2)** from the other remaining questions.
2. All questions carry equal weight.
3. Credit will be given for orderly presentation of **relevant** materials.

Question 1 (COMPULSORY)

(a)

(i) Discuss stationarity versus nonstationarity in panel data econometrics. **[2 marks]**

(ii) Given a typical panel data model, how would you test for the ADF unit roots in it.[Outline the steps involved in the ADF test] **[4 marks]**

(iii) What is panel cointegration? **[2 marks]**

(iv) Consider a panel regression of the form:

$$y_{it} = \alpha_i + \beta x_{it} + e_{it}$$

$$y_{it} = y_{it-1} + u_{it}$$

$$x_{it} = y_{it-1} + \varepsilon_{it}$$

t = 1,2,...,T; i = 1,2,...,N

How would you conduct the KAO [Engle-Granger Based] cointegration test? **[3 marks]**

(b)

(i) What is the relationship between the hazard rate and the survivor rate functions [Mathematical derivation is important] **[6 marks]**

(ii) Given the basic Cox proportional hazard model

$$\mu(t|x_i) = \mu_0(t)exp(\beta_1 x_1 + \beta_2 x_2 +, \dots, \beta_k x_k)$$

Why is this model considered as a semi-parametric model? Discuss the method of estimation of the parameters $\beta_1,\beta_2,\dots,\beta_k$ **[3 marks]**

Question 2

Consider the following model:

$$y_{it} = \alpha_i + \beta x_{it} + u_{it}, \qquad t = 1, \dots, T$$

The panel is balanced. There are N individuals. u_{it} is an idiosyncratic error term with the usual properties. a_i is the individual effect that does not vary overtime. y_{it} is the dependent variable and x_{it} is a scalar covariate.

(a) Write down formulae for the pooled OLS, the Fixed Effects, and the between estimators of β. Denote these as $\widehat{\beta}$, $\hat{\beta}_\omega$ and $\hat{\beta}_b$ respectively. **[6 marks]**

(b) In the Stata output below someone is trying to compute estimates for the wage- gender relationship using the data taken from the 1991 and 1997 Waves of the British Household Panel Survey. The dependent variable is the log nominal weekly earnings (*lw*), personal identifier (pid) and the independent variable is a dummy for whether the individual is male (*m*). The sample size is 1685. The three 2 by 2 matrices report sample variance and covariances. Complete this person's calculations. Comment on your result. **[8 marks]**

```
. by pid: egen lw1=mean(lw)

. by pid: egen m1=mean(m)

. gen lw2=lw-lw1

. gen m2=m-m1

. li pid yr lw m lw1 m1 lw2 m2 in 1/20
```

	pid	yr	lw	m	lw1	m1	lw2	m2
1.	10023526	1	5.273769	0	5.228222	0	.045547	0
2.	10023526	7	5.182675	0	5.228222	0	-.0455465	0
3.	10028005	1	5.848423	1	5.88939	1	-.0409675	0
4.	10028005	7	5.930357	1	5.88939	1	.040967	0
5.	10060111	1	5.790793	1	5.908262	1	-.1174688	0
6.	10060111	7	6.025732	1	5.908262	1	.1174693	0
7.	10061649	1	6.06943	1	6.117547	1	-.0481167	0
8.	10061649	7	6.165663	1	6.117547	1	.0481162	0
9.	10071687	1	5.481409	1	5.517569	1	-.03616	0
10.	10071687	7	5.553729	1	5.517569	1	.0361605	0
11.	10071717	1	5.394397	0	5.549475	0	-.1550775	0
12.	10071717	7	5.704552	0	5.549475	0	.1550775	0
13.	10080643	1	5.848423	0	5.701295	0	.1471276	0
14.	10080643	7	5.554167	0	5.701295	0	-.1471281	0
15.	10092986	1	5.442957	0	5.686657	0	-.2436996	0
16.	10092986	7	5.930357	0	5.686657	0	.2437	0
17.	10094083	1	6.620843	1	6.529266	1	.0915766	0
18.	10094083	7	6.437689	1	6.529266	1	-.0915771	0
19.	10127666	1	5.848423	0	5.949624	0	-.1012015	0
20.	10127666	7	6.050826	0	5.949624	0	.101202	0

```
. corr lw m, cov
(obs=3370)

             |       lw        m
-------------+------------------
          lw |  .282251
           m |  .071346  .233713


. corr lw1 m1, cov
(obs=3370)

             |      lw1       m1
-------------+------------------
         lw1 |  .221596
          m1 |  .071346  .233713


. corr lw2 m2, cov
(obs=3370)

             |      lw2       m2
-------------+------------------
         lw2 |  .060655
          m2 |        0        0
```

(c) The random effects estimator is found by running OLS on the following transformed equation:

$$y_{it} - \mu\bar{y}_i = (1-\mu)\beta_0 + \beta(x_{it} - \mu\bar{x}_i) + (v_{it} - \mu\bar{v}_i),$$

where

$$\mu = 1 - \sqrt{\sigma_u^2 \big/ (\sigma_u^2 + T\sigma_a^2)} = 1 - \sqrt{\theta},$$

and σ_u^2 and σ_a^2 are the population variances of u_{it} and a_i respectively.

(i) Clearly $0 \leq \mu \leq 1$. Explain what happens when $\mu = 0$ and when $\mu = 1$.

(ii) The person above discovers the Stata command for the Random Effects estimator, and generates the output below. Comment on the relationship between the estimate on the male dummy, and how close/far away from the three estimates computed in (b) above **[6 marks]**

```
. xtreg lw m, re i(pid)

Random-effects GLS regression                   Number of obs      =      3370
Group variable (i): pid                         Number of groups   =      1685

R-sq:  within  =      .                         Obs per group: min =         2
       between = 0.0983                                        avg =       2.0
       overall = 0.0772                                        max =         2

Random effects u_i ~ Gaussian                   Wald chi2(1)       =    183.45
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
          lw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           m |   .3052729   .0225388    13.54   0.000     .2610976    .3494482
       _cons |   5.437063   .0178597   304.43   0.000     5.402059    5.472067
-------------+----------------------------------------------------------------
     sigma_u |  .37330476
     sigma_e |  .34824533
         rho |  .53468816   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Question 3

We are interested in explaining unemployment. We believe that the probability of being unemployed is affected by gender, age and education level. Let y_i denote unemployment (y_i=1 if a person i is unemployed and 0 otherwise), let x_{1i} denote gender (x_{1i}=1 if person i is male and 0 otherwise, let x_{2i} denote age (x_{2i} is the age of person i, continuously measured) and let x_{3i}denote education level (x_{3i}= 1 if person i has a university degree and otherwise). We have data on these variables for n = 7867 individuals taken from Kenya Household Panel Survey. For the purpose of explaining the probability of being unemployed, we consider a binary choice model:

$$y_i^* = x_i'\beta + \varepsilon_{i,} \quad \varepsilon_i \text{ iid}, \qquad i = 1,2, \dots, n$$

$$y_i = \begin{cases} 1 & y_i^* \geq 0 \\ 0 & y_i^* < 0 \end{cases}$$

Let $F(.)$ denote the cdf of ε,

(a) Give an expression for log likelihood function of the logit model. **[5 marks]**

(b) The binary choice model is now estimated both by logit and probit (see stata output in appendix A). How do you interpret the coefficient estimates of β from the logit and probit? **[4 marks]**

(c) Using the logit estimates, what is the difference in the predicted probabilities of being unemployed between men and women? **[6 marks]**

(d) How much does the probability of being unemployed increase/decrease with age. Comment on your findings? **[5 marks]**

Question 4

The Belgian government is contemplating increasing the tax on tobacco in order to lower the incidence of smoking. A consultant to the government therefore wishes to examine whether such a tax increase will actually lower the demand for tobacco. He knows that in order to assess this he needs to take both the price and income effect into account. He has an estimate of the price elasticity for tobacco and he turns to you for assistance in estimating the income elasticity. The data he has available for estimating the income elasticity is the Belgian Household Budget Survey 1995-1996, which contains information on household, how many adults live in the household as well as the age class of the head of the household for $n = 2724$ households. He has estimated a linear regression model on this data by running OLS of the budget share for tobacco (*btobacco*) on log total expenditure (*lnx*), number of children in the household (*nkids, nkids2*), number of adults in the household (*nadults*) and age (*age*). The stata output for this regression is given in appendix B.Denoting the budget share of tobacco by ω and the log total expenditure by $\ln x$, the income elasticity e for tobacco can be calculated by the formula

$$e = \frac{1}{\omega}\frac{\partial \omega}{\partial lnx} + 1$$

(a) Calculate the income elasticity e_{OLS} resulting from the OLS. **[6 marks]**

(b) As an alternative to the OLS, you suggest to model tobacco expenditures by a tobit model. Why would a tobit model be more appropriate for analyzing this data? Write down the appropriate tobit model. **[5 marks]**

(c) The Stata output for the estimation of the tobit model is given in appendix B. Calculate the income elasticity e_{TOBIT} resulting from the tobit. Comment on your findings for e_{OLS} and e_{TOBIT}. **[9 marks]**

Question 5

There are many factors that influence elections. One such factor that has received considerable attention is the impact of campaign expenditures on election outcomes. The following equation describes the percentage of the vote (pctvote) received by a candidate (measured on a 0% - 100% scale):

$$pctvote = \beta_0 + \beta_1 \log(exp\_cond) + \beta_2 \log(\exp\_opp) + \beta_3 Party + u$$

where log(exp_cand) is the log of the candidate's own expenditures, log(exp_opp) is the log of the candidate's opponent's expenditures (with expenditures measured in thousands of dollars), and Party is the political party of the candidate (1 if Democrat, 0 if Republican).

(a) Using data on 173 congressional races for the U.S. House of Representatives in the 1992 election, the following equation was estimated:

$$\widehat{pctvote} = 51.13 + 6.30 \log(exp\_cond) - 6.67\beta_2 \log(\exp\_opp) + 1.21 Party$$

$R^2 = 0{:}786$

SSR = 10351.2

How do you interpret the coefficients on log (exp_cand) and log (exp_opp)? How do you interpret the coefficient on Party? **[5 marks]**

(b) The following standard errors (and covariance) were obtained from the computer regression output:

Estimated Standard errors	
S.E.(β_0)	2.90
S.E.(β_1)	0.37
S.E.(β_2)	0.39
S.E.(β_3)	1.34
Cov(β_1,β_2)	-0.00057

Test the hypothesis that political party has no effect on the percentage of the vote a candidate receives. (Notes: Be sure to state the null and alternative hypotheses. Use a 5% significance level for a two-sided test. As the degrees of freedom are greater than 150, you can use the critical values from a standard normal distribution (i.e., ±1.96). You can test the hypothesis using either a t-statistic or a confidence interval.) **[5 marks]**

(c) It appears from the estimated regression that the coefficients on log(exp_cand) and log(exp_opp) are of equal magnitudes and opposite signs. Test the hypothesis H_0: $\beta_1 = -\beta_2$. (Note: This hypothesis also means that it is only the difference in log expenditures between the two candidates that matters.) **[5 marks]**

(d) Your colleague does not believe that campaign expenditures matter in an election, and therefore estimates the following regression:

$\hat{pctvote} = 45.70 + 8.65 Party$
$R^2 = 0:066$
$SSR = 45258.1$

Test the joint hypothesis that campaign expenditures do not matter (H_0: $\beta_1 = \beta_2 = 0$). (The critical value for the test is 3.02.) **[5 marks]**

APPENDICES

Appendix A

```
   obs:          7,867                                  BHPS annual panel 1991-2002
  vars:              4                                  13 Oct 2005 07:21
  size:         86,537 (99.8% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
age             byte   %8.0g       aage       age at date of interview
unemp           byte   %9.0g                  unemployed
male            byte   %9.0g                  =1 if individual is male
degree          float  %9.0g                  =1 if individual has university
                                                degree
-------------------------------------------------------------------------------
```

```
 summarize male age degree

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       unemp |      7867    .0526249    .2232977          0          1
        male |      7867    .5215457    .4995673          0          1
         age |      7867    38.84912    11.85259         16         65
      degree |      7867    .1552053    .3621233          0          1
```

```
. logit unemp male age degree

Iteration 0:   log likelihood = -1621.9607
Iteration 1:   log likelihood = -1574.1991
Iteration 2:   log likelihood = -1571.2739
Iteration 3:   log likelihood = -1571.2507
Iteration 4:   log likelihood = -1571.2507

Logistic regression                               Number of obs   =       7867
                                                  LR chi2(3)      =     101.42
                                                  Prob > chi2     =     0.0000
Log likelihood = -1571.2507                       Pseudo R2       =     0.0313

------------------------------------------------------------------------------
       unemp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   .2148666   .1025754     2.09   0.036     .0138225    .4159107
         age |  -.0361777   .0044211    -8.18   0.000    -.0448429   -.0275125

      degree |  -.9472709   .1926279    -4.92   0.000    -1.324815   -.5697272
       _cons |  -1.574088   .1721039    -9.15   0.000    -1.911406   -1.236771
------------------------------------------------------------------------------

. probit unemp male age degree

Iteration 0:   log likelihood = -1621.9607
Iteration 1:   log likelihood = -1574.2058
Iteration 2:   log likelihood = -1573.3143
Iteration 3:   log likelihood = -1573.3125

Probit regression                                 Number of obs   =       7867
                                                  LR chi2(3)      =      97.30
                                                  Prob > chi2     =     0.0000
Log likelihood = -1573.3125                       Pseudo R2       =     0.0300

------------------------------------------------------------------------------
       unemp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   .0960036   .0479677     2.00   0.045     .0019885    .1900186
         age |  -.0159835   .0019968    -8.00   0.000    -.0198971   -.0120698
      degree |  -.4218795   .0816118    -5.17   0.000    -.5818358   -.2619233
       _cons |  -1.028603   .0810051   -12.70   0.000     -1.18737   -.8698358
------------------------------------------------------------------------------
```

AERC
CREA

Appendix B

```
. reg btobacco lnx age nadults nkids2 nkids

      Source |       SS       df       MS              Number of obs =    2724
-------------+------------------------------           F(  5,  2718) =   40.32
       Model |  .116758246     5  .023351649           Prob > F      =  0.0000
    Residual |   1.5741102  2718  .000579143           R-squared     =  0.0691
-------------+------------------------------           Adj R-squared =  0.0673
       Total |  1.69086845  2723  .000620958           Root MSE      =  .02407

------------------------------------------------------------------------------
    btobacco |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnx |  -.0141745   .0011452   -12.38   0.000    -.0164202   -.0119289
         age |  -.0025072   .0003865    -6.49   0.000    -.0032651   -.0017493
     nadults |   .0027508   .0006524     4.22   0.000     .0014716    .0040301
      nkids2 |  -.0047776   .0022332    -2.14   0.032    -.0091565   -.0003987
       nkids |    .001168   .0005623     2.08   0.038     .0000654    .0022705
       _cons |   .2069766   .0151311    13.68   0.000     .1773069    .2366462
------------------------------------------------------------------------------

. predict wreg, xb

. summarize wreg

    Variable |     Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------
        wreg |    2724    .0321908    .0062637   .00128599    .060071
```

```
. tab d2

 dummy=1 if |
    tobacco |
expenditure |
         >0 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      1,688       61.97       61.97
          1 |      1,036       38.03      100.00
------------+-----------------------------------
      Total |      2,724      100.00
```

```
. tobit btobacco lnx age nadults nkids2 nkids , ll(0)

Tobit regression                                  Number of obs   =       2724
                                                  LR chi2(5)      =     145.58
                                                  Prob > chi2     =     0.0000
Log likelihood =  746.40082                       Pseudo R2       =    -0.1081

------------------------------------------------------------------------------
    btobacco |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnx |  -.0256124   .0027221    -9.41   0.000      -.03095   -.0202748
         age |   -.006387   .0009186    -6.95   0.000    -.0081882   -.0045858
     nadults |   .0076941    .001545     4.98   0.000     .0046645    .0107237
      nkids2 |  -.0135256   .0054335    -2.49   0.013    -.0241798   -.0028714
       nkids |   .0029758   .0012966     2.29   0.022     .0004333    .0055183
       _cons |    .334203   .0357935     9.34   0.000     .2640178    .4043883
-------------+----------------------------------------------------------------
      /sigma |   .0483493   .0011926                      .0460108    .0506879
------------------------------------------------------------------------------
  Obs. summary:       1688  left-censored observations at btobacco<=0
                      1036     uncensored observations
                         0 right-censored observations
```

```
. predict wtobit, e(0,1)

. summarize wtobit

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      wtobit |      2724    .0334894    .0040911   .0230732   .0575392
```