AERC CREA

AFRICAN ECONOMIC RESEARCH CONSORTIUM

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

JOINT FACILITY FOR ELECTIVES (JFE) 2011

JUNE – SEPTEMBER

ECONOMETRICS THEORY AND PRACTICE II

Second Semester: Final Examination

Duration: 3 Hours **Date: Thursday, September 22, 2011**

INSTRUCTIONS:

1. This examination has **FIVE (5) QUESTIONS**. Each question carries **TWENTY (20)** marks.
2. You are required to attempt **ANY THREE (3) QUESTIONS**.
3. Relevant formulae are embedded in the questions wherever they are necessary.
4. Show your derivations and mathematical steps in detail.
5. You may use an unprogrammable calculator.

Question 1

Suppose you are a micro-econometrician hired by the African Econometrics Society (AES) to analyse the factors that determined the choice of econometrics by JFE alumni. You sent questionnaires to 1000 CMAP JFE alumni over the period 2004-2010 and there was 100% response rate. The following is a portion of the questionnaire.

No.	symbol	Question	Response
1	y	*Did you do econometrics theory and practice at the JFE?*	1: Yes 0: No
2	x^a	*What was your age at the time of entering the CMAP JFE?*	State approximate years
3	x^g	*What is your gender?*	0: Female 1: Male

You are interested in predicting the probability that a CMAP JFE student will select econometrics theory and practice as one of the electives conditional on the age and gender i.e. $\Pr(y_i = 1 | x_i^a, x_i^g)$.

(a) Assume that the binary variable y_i follows a linear probability model (LPM) for the 1000 observations ($i = 1,2,..,1000$)

(i) State the linear probability model for the choice of econometrics. **(1 Mark)**

(ii) Prove that the LPM model is inherently heteroscedastic. **(9 Marks)**

(b) Assume that the utility that a CMAP JFE student gets by selecting econometrics theory and practice can be represented as $U_i = \beta_0 + \beta_1 x_i^a + \beta_2 x_i^g + \varepsilon_i$ and ε_i follows a logistics distribution

(i) Derive the log-likelihood function. **(7 Marks)**

(ii) State the steps that one can follow to find the maximum likelihood estimates (MLE) using the log-likelihood function above. (You do not need to derive the estimates). **(3 Marks)**

Question 2

Consider the following binary choice model

$$\Pr(y = 1|income_i, gender_i) = \Phi(\beta_0 + \beta_1 income_i + \beta_2 gender_i + \beta_3 income_i^2)$$

Where Φ is the cumulative density function for normal distribution

(a) Find the partial effects of income on the response probability and comment on the nature of the result. **(5 Marks)**

(b) How would you estimate the partial effects of income on the response probability? **(5 Marks)**

(c) How do you estimate the partial effect of gender on the response probability? **(5 Marks)**

(d) Suppose you are interested in finding the probability that a student's grade would improve (grade=1) conditional on grade point average (gpa) and gender. You estimate the results and the stata output is shown below:

```
Probit regression                               Number of obs   =         32
                                                LR chi2(2)      =      15.15
                                                Prob > chi2     =     0.0005
Log likelihood = -13.016522                     Pseudo R2       =     0.3679

------------------------------------------------------------------------------
       grade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |   1.784757    .659368     2.71   0.007     .4924196    3.077095
      gender |   1.421516   .5860173     2.43   0.015     .2729431    2.570089
       _cons |   -6.78252   2.233819    -3.04   0.002    -11.16072   -2.404315
------------------------------------------------------------------------------


Marginal effects after probit
      y  = Pr(grade) (predict)
         =  .27519213
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     gpa |   .5957281      .22104    2.70   0.007   .162502  1.02895   3.11719
 gender*|   .4688015      .17024    2.75   0.006   .135145  .802458     .4375
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Interpret the marginal effects. **(5 Marks)**

Question 3

(a) (i) Distinguish between multinomial logit and conditional logit models in terms of assumptions about the distribution function of the error components, IIA assumption, specification of the dependent variable in stata econometrics software and nature of regressors. **(5 Marks)**

(ii) Explain how and to what extent the nested logit and multinomial probit models address the IIA problem. **(5 Marks)**

(iii) What is an inclusive value parameter in a nested logit model? State the null and alternative hypotheses for a likelihood ratio test for the IIA using the inclusive parameter value. **(5 Marks)**

(b) Suppose you are a health economist interested in the health of residents in a particular region. You have collected data and run a multinomial logit regression model. The stata output is as follows:

```
mlogit health age sex height,base(1) nolog

Multinomial logistic regression                   Number of obs   =      10335
                                                  LR chi2(12)     =    1654.77
                                                  Prob > chi2     =     0.0000
Log likelihood = -14937.014                       Pseudo R2       =     0.0525

------------------------------------------------------------------------------
      health |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
fair         |
         age |  -.0194561   .0035064    -5.55   0.000    -.0263285   -.0125838
         sex |   .3884021   .1268433     3.06   0.002     .1397937    .6370105
      height |   .0013626    .006785     0.20   0.841    -.0119357    .0146608
       _cons |   1.145548   1.333744     0.86   0.390    -1.468542    3.759637
-------------+----------------------------------------------------------------
average      |
         age |  -.0478771   .0032818   -14.59   0.000    -.0543092    -.041445
         sex |   .3397713   .1196164     2.84   0.005     .1053275    .5742151
      height |   .0049578   .0063937     0.78   0.438    -.0075736    .0174893
       _cons |   2.692775    1.25598     2.14   0.032     .2311001     5.15445
-------------+----------------------------------------------------------------
good         |
         age |   -.065821   .0033284   -19.78   0.000    -.0723446   -.0592974
         sex |   .5617364   .1234116     4.55   0.000     .3198541    .8036187
      height |   .0252566   .0065864     3.83   0.000     .0123475    .0381657
       _cons |  -.3449937   1.293057    -0.27   0.790    -2.879339    2.189352
-------------+----------------------------------------------------------------
excellent    |
         age |  -.0774643   .0033786   -22.93   0.000    -.0840862   -.0708425
         sex |   .4733492   .1258432     3.76   0.000     .2267011    .7199973
      height |    .034241   .0067113     5.10   0.000      .021087    .0473949
       _cons |  -1.318139    1.31682    -1.00   0.317     -3.89906    1.262781
------------------------------------------------------------------------------
(health==poor is the base outcome)
```

The marginal effects for good health is

```
Marginal effects after mlogit
      y  = Pr(health==4) (predict, outcome(4))
         =  .26459583
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     age |  -.0035055      .00027  -12.96   0.000  -.004036 -.002975   47.5658
   sex1* |  -.0379626       .0127   -2.99   0.003  -.062849 -.013076   .474988
  height |   .0024057      .00068    3.56   0.000   .001082  .003729   167.653
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0(female) to 1(male)
```

Height is in cm

Interpret the marginal effects. **(5 Marks)**

AERC CREA

Question 4

(a) (i) Distinguish between censoring and truncation of data. **(4 Marks)**

(ii) If you compare "censoring" and "truncation", which of these two entails more information loss? **(1 Mark)**

(iii) Briefly explain the key limitations of the Tobit 1 model **(5 Marks)**

(b) Consider the latent variable model

$$y_1^* = x_1'\beta_1 + \varepsilon_1$$

$$y_2^* = x_2'\beta_1 + \varepsilon_2$$

But we observe

$$y_1 = \begin{cases} 1 & if \quad y_1^* > 0 \\ 0 & if \quad y_1^* \le 0 \end{cases}$$

$$y_2 = \begin{cases} y_2^* & if \quad y_1^* > 0 \\ - & if \quad y_1^* \le 0 \end{cases}$$

Where the symbol – means missing

(i) Find a general expression for $E[y_2|x_1, x_2, y_1^* > 0]$ **(3 Marks)**

(ii) Assume that ε_1 and ε_2 are related as follows

$$\varepsilon_2 = \sigma_{12}\varepsilon_1 + \eta_t$$

Where η_t is independent of ε_1 and has mean zero.

Prove that the general expression in Question 4(b)(i) above can be written as

$$E[y_2|x_1, x_2, y_1^* > 0] = x_2'\beta_2 + \sigma_{12}E[\varepsilon_1|x_1, x_2, \varepsilon_1 > -x_1'\beta_1]$$ **(7 Marks)**

Question 5

(a) State the advantages of using panel data. **(5 Marks)**

(b) Consider a one-way error component linear panel data model

$$y_{it} = \alpha + \mu_i + x_{it}'\beta + v_{it} \qquad i = 1,2,\dots,N \text{ and } t = 1,2,\dots,T$$

$$v_{it} \sim Niid(0, \sigma_v^2) \text{ and } N \to \infty \text{ while T is small.}$$

What criteria (including test) would you use to decide between fixed effects (FEM) model and random effects model (REM)? **(5 Marks)**

(c) Derive the WITHIN estimator for the model in 5(b) above. **(5 Marks)**

(d) Suppose the model in 5(b) is re-specified as a two-way error component:

$$y_{it} = \alpha + \mu_i + \lambda_t + x'_{it}\beta + v_{it}$$

Where μ_i are cross-section specific-fixed effects

λ_t are time-specific fixed effects

You have data for twelve (12) African countries on real private consumption (RCONS) and real GDP (RGDP) over the period 1997 to 2010. You estimate the model in Eviews econometrics software and find the following results.

Dependent Variable: LN_RCONS
Method: Pooled Least Squares
Sample: 1997 2010
Included observations: 14
Cross-sections included: 12
Total pool (unbalanced) observations: 163

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.063740	0.633160	7.997570	0.0000
LN_RGDP	0.498205	0.059060	8.435613	0.0000
Cross-section fixed effects (μ_i)				
Botswana	-1.244337			
Burkina Faso	1.731393			
Burundi	1.943289			
Kenya	1.347627			
Madagascar	-0.200967			
Mauritius	0.214007			
Morocco	-2.576373			
Nigeria	-1.459456			
Rwanda	-1.911089			
Sierra Leone	1.882629			
South Africa	1.205361			
Tanzania	-0.932085			
Time fixed effects (λ_i)				
1997	-0.003013			
1998	0.023721			
1999	0.031312			
2000	0.025417			
2001	0.019385			
2002	0.000829			
2003	0.024241			
2004	0.055336			
2005	0.082757			
2006	0.057133			
2007	0.064770			
2008	0.120360			
2009	0.171388			

2010	0.186043

Effects Specification

Cross-section fixed (dummy variables)
Period fixed (dummy variables)

R-squared	0.999027	Mean dependent var	10.40442
Adjusted R-squared	0.998849	S.D. dependent var	2.939862
S.E. of regression	0.099736	Akaike info criterion	-1.627333
Sum squared resid	1.362774	Schwarz criterion	-1.133851
Log likelihood	158.6276	F-statistic	5624.744
Durbin-Watson stat	0.945732	Prob(F-statistic)	0.000000

(i) Interpret the country-specific fixed effects (Botswana and Kenya only) and the time-specific fixed effects (1997 and 2010 only). **(4 Marks)**

(ii) Suppose the model is re-specified as a dynamic panel data. What is the consequence of the re-specification on a fixed effects model and random effects model? **(1 Mark)**