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

JOINT FACILITY FOR ELECTIVES

JUNE – OCTOBER 2008

ECONOMETRICS THEORY & PRACTICE II

Second Semester: Final Examination

Time: 09.00 AM – 12.00 Noon **Date: Tuesday, September 29, 2008**

INSTRUCTIONS:

(i) Choose **THREE (3)** out of the following **FIVE (5)** questions. All questions have equal weight.

(ii) You can use unprogrammable calculators.

(iii) Relevant formulae are embedded in the questions wherever they are necessary.

(iv) Show your derivations and mathematical steps in detail.

Question 1

1.1 Suppose you are the micro-econometrician for ABC Research International Plc. You carried a survey on 4010 small and medium enterprises (SMEs) in country X. The following is a portion of the questionnaire.

No.	**Question**	**Response**
$y =$	*Have you ever received credit to operate or expand your business?*	0: No 1: Yes
$x =$	*How many people does your enterprise hire?*	State number of employees
$d =$	*What is the gender of the owner of the SME?*	0: Female 1: Male

You are interested in predicting the probability that an SME will receive credit to operate or expand the business given x and d i.e. $\Pr\left(y_i = 1 | x, d\right)$.

Assume that the binary variable y_i follows a linear probability model (LPM)

(a) Derive the log-likelihood function (assume no constant in your specification) **(25 points)**

(b) State what condition must hold for the log-likelihood function in (a) to be well defined? **5 points)**

(c) Using the results from (a) explain why the LPM model is inherently heteroscedastic. **(10 points)**

1.2 Assume that there is an underlying latent variable y_i^* that drives the decision to either receive credit or not in question (1.1). The latent variable has the following specification

$$y_i^* = \beta_0 + \beta_1 x_i + \beta_2 d_i + u_i$$

where the errors u_i are logistic distributed with the density $f(u) = \frac{e^{-u}}{(1+e^{-u})^2}$ and cdf of

$$F(u) = \frac{e^u}{1+e^u}.$$

We observe an SME getting credit as $y_i = \begin{cases} 1 & if \quad y_i^* > 0 \\ 0 & if \quad y_i^* \le 0 \end{cases}$

(a) Write down the log-likelihood function **(30 points)**

(b) Explain the steps for computing maximum likelihood estimates for β_1 and β_2 from the log-likelihood function in question 1.2 (a). You are not expected to do the actual computations. **(10 points)**

(c) You estimate the logit model and find the following results

```
Dependent variable is receive
 Logistic regression                                 Number of obs   =       4010
                                                     LR chi2(2)      =     100.01
                                                     Prob > chi2     =     0.0000
 Log likelihood =  -1781.554                         Pseudo R2       =     0.0273

 ------------------------------------------------------------------------------
    receive |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
 -----------+------------------------------------------------------------------
 employment |   .0111045   .0027068     4.10   0.000     .0057992    .0164098
     gender |   .1851423   .0847502     2.18   0.029     .0190349    .3512496
      _cons | -1.715906    .0626053   -27.41   0.000     -1.83861   -1.593202
  ------------------------------------------------------------------------------
```

Provide a semi-elasticity interpretation of the coefficients for employment and gender. **10 points)**

(d) There has been a general public complaint that credit providers engage in gender discrimination and failure to nurture small SMEs. The logit estimation results (above) by ABC Research International Plc are to be used to make policy decision regarding the public complaint. You present the results in question (1.2c) and a semi-elasticity interpretation to your managing director, Mrs. Mkopo. She suggests that estimates of the marginal effects would be more informative than the semi-elasticities. You compute the marginal effects and find the following.

```
Marginal effects after logit
y  = Pr(receive) (predict)
   =  .17135018
------------------------------------------------------------------------------
variable  |    dy/dx      Std. Err.     z    P>|z|   [    95% C.I.   ]      X
 ---------+--------------------------------------------------------------------
employment|   .0015767      .00039    4.06   0.000   .000816  .002337   4.12372
    gender|   .0262687      .01199    2.19   0.028   .002766  .049772   .507857
 ------------------------------------------------------------------------------

X-Mean values
```

Using the marginal effects above.

(i) Comment on whether credit providers prefer large SMEs (those which employ more people) to the small ones. **(5 points)**

(ii) Have credit providers been engaging in gender discrimination? **(5 points)**

Question 2

2.1 Suppose a household in Arusha has three mutually exclusive options to take when a member of the family is sick: nearby clinic (C), nearby private hospital (PR) and go to a public hospital (P). The utility for outcome j for individual i is given by $U_{ij} = \beta_1 age_i + \beta_2 income_i + \varepsilon_{ij}$

Where

ε_{ij} is iid extreme value distributed

age_i is the age of the sick person

$income_i$ is the income of the head of the household

Suppose what is observed is $y_i = \begin{cases} 1 & if \quad y_i = j \\ 0 & if \quad y_i \neq j \end{cases}$

Where $j = C, PR, P$.

(a) Assume that the reference category is C. State the conditional probability for C, PR and P

i.e. $\Pr\left[y_{iC} = C \middle| age_i, income_i\right]$, $\Pr\left[y_{iPR} = PR \middle| age_i, income_i\right]$ and $\Pr\left[y_{iP} = P \middle| age_i, income_i\right]$

(20 points)

(b) What property of multinomial logit models informs the identification process of setting one of the outcomes as a reference category? **(10 points)**

(c) What is the consequence of not setting one alternative as a reference category? **(10 points)**

(d) Assume that the public hospital (P) in Arusha is modernised by the government to have state-of-the art facilities. Show mathematically how the modernisation of the public hospital would affect the decision of a household debating between going to a nearby clinic (C) or nearby private hospital (PR) under multinomial logit model. **(15 points)**

2.2 Suppose a researcher has prepared a questionnaire to administer on the households in Arusha. The questionnaire has questions on *age*, *income* and *choice of health facility (i.e.* whether C, PR and P*)*. The researcher is very excited about McFadden's Conditional logit model and would like to apply it on the health facility choice problem in 2.1.

(a) Is the questionnaire sufficient to collect information to estimate a conditional logit model? **(10 points)**

(b) If not, suggest other variables that the researcher should collect. **(10 points)**

2.3 The inclusive value parameter (logsum) in a nested logit model plays an important role in the likelihood ratio test for the independence of irrelevant alternatives (IIA). State the null and alternative hypotheses for this test for the C, PR, and P alternatives. **(10 points)**

2.4 Assume that the utilities (U_j) for the different health facilities can be described as follows.

$U_c = V_C + \varepsilon_C$ for the clinic

$U_{PR} = V_{\Pr} + \varepsilon_{\Pr}$ for private hospital

$U_P = V_P + \varepsilon_P$ for the public hospital

Where V_j is the deterministic part of the utility

ε_j is the error term

Use the above information to explain how the multinomial probit model remedies the IIA assumption problem. **(15 points)**

Question 3

3.1 Consider the latent variable model

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

$$y_2^* = x_2'\beta_2 + \varepsilon_2$$

But we observe

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

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

(a) Find a general expression for $E\left[y_2 \middle| x, y_1^* > 0\right]$ **.** Where x is composed of distinct elements of x_1 and x_2 i.e. $x = (x_1, x_2)$**.** **(15 points)**

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

$$\varepsilon_2 = \theta\varepsilon_1 + \eta_t$$

Where η_t is independent of ε_1 and has mean zero. Specialise the general expression in 3(a). **(30 points)**

(c) Assume that $\varepsilon_1 \sim N(0,1)$. Prove that $E\left[y_2 \middle| x, y_1^* > 0\right] = x_2'\beta_2 + \sigma_{12}\lambda(x_1'\beta_1)$ Where $x = (x_1, x_2)$**.** (Hint: for $z \sim N(0,1)$, $E\left[z \middle| z > c\right] = \dfrac{\phi(c)}{1-\Phi(c)} = \dfrac{\phi(-c)}{\Phi(-c)}$).

(30 points)

3.2 Assume that you have data on y_1, y_2, x_1 and x_2. Explain how you would compute the estimates for β_1 and β_2 with an econometric software that has probit and OLS commands only. **(25 points)**

Question 4

4.1 Consider a one-way error component linear panel data model

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

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

What criteria (including test) would you use to decide between fixed effects and random effects model? **(10 points)**

4.2 Using the panel data model in 4.1 above derive the *WITHIN* estimator. **(25 points)**

4.3 Explain the Hausman specification test for random effect model vs. fixed effects model. **(20 points)**

4.4 Suppose the model in 4.1 is re-specified as a two-way error component model as follows

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

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 consumption (RCONS) and GDP(RGDP) over the period 1990 to 2003. You estimate the model in Eviews econometrics software and find the following results.

Dependent Variable: LN_RCONS
Method: Pooled Least Squares
Sample: 1990 2003
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)				
1990	-0.003013			

1991	0.023721
1992	0.031312
1993	0.025417
1994	0.019385
1995	0.000829
1996	0.024241
1997	0.055336
1998	0.082757
1999	0.057133
2000	0.064770
2001	0.120360
2002	0.171388
2003	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

Interpret the country-specific fixed effects (**Botswana** and **Kenya only**) and the time-specific fixed effects (**1990** and **2003 only**). **(10 points)**

4.5 Suppose the panel data model was re-specified in a dynamic form as follows

$$y_{it} = \alpha + \mu_i + \delta y_{it-1} + x'_{it}\beta + \nu_{it}$$

(a) Explain the Nickell (1981) bias in WITHIN estimation of the above dynamic panel data model. **(15 Points)**

(b) State two GMM methods that can be used to deal with the problem. **(10 points)**

4.6 You conducted panel unit root test for the consumption variable and the summary of unit root tests statistics are presented below.

Pool unit root test: Summary
Sample: 1990 2003
Series: LN_RCONS_BOTS, LN_RCONS_BURKF, LN_RCONS_BUR, LN_RCONS_KEN, LN_RCONS_MADAG, LN_RCONS_MAURIT, LN_RCONS_MOR, LN_RCONS_NIG, LN_RCONS_RWA, LN_RCONS_SIERL, LN_RCONS_RSA, LN_RCONS_TAN
Exogenous variables: Individual effects
Automatic selection of maximum lags
Automatic selection of lags based on SIC: 0 to 2
Newey-West bandwidth selection using Bartlett kernel

Method	Statistic	Prob.**	Cross-sections	Obs
Levin, Lin & Chu t*	-0.10000	0.4602	12	147
Im, Pesaran and Shin W-stat	2.13041	0.9834	12	147
ADF - Fisher Chi-square	17.7157	0.8166	12	147
PP - Fisher Chi-square	15.2141	0.9143	12	156
Hadri Z-stat	6.89951	0.0000	12	168

** Probabilities for Fisher tests are computed using an asympotic Chi -square distribution. All other tests assume asymptotic normality.

Interpret the Im, Pesaran and Shin (IPS)-W test and Hadri Z-statistic for panel unit root tests. **(10 points)**

Question 5

5.1 Define the following concepts

(a) Failure function **(5 points)**

(b) Survivor function **(5 points)**

(c) Hazard function **(5 points)**

5.2 **(a)** Show that the hazard rate function, $\lambda(t)$, is related to the survivor function, $S(t)$, as follows $\lambda(t) = -\frac{\partial \ln S(t)}{dt}$. **(35 points)**

(b) Hence or otherwise prove that the survivor function can be represented as

$$S(t) = \exp\left(-\int_0^t \lambda(u)du\right).$$ **(30 points)**

5.3 You have duration data for 3343 individuals on the following

Spells: period of unemployment

Unemployment claim: Unemployment claim insurance (1 if filed claim)

Wage: Weekly earnings

Age: Age of the unemployed person

You fit the following Cox proportional hazards model on the spells.

$$\lambda(t|uc, \ln wage, age) = \lambda_0(t)\exp(\beta_1 uc + \beta_2 \ln wage + \beta_3 age)$$

The results are as follows.

```
failure _d:  1 (meaning all fail)
analysis time _t:  spell

 Cox regression -- Breslow method for ties

     No. of subjects =         3343                   Number of obs   =      3343
     No. of failures =         3343
     Time at risk    =        20887
                                                       LR chi2(3)      =    372.93
     Log likelihood  =   -23921.542                    Prob > chi2     =    0.0000

  ------------------------------------------------------------------------------
              _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
   -----------+----------------------------------------------------------------
Unemployment claim |   .5150703   .0188909   -18.09   0.000     .4793443    .5534591
            lnwage |   1.122673   .0372087     3.49   0.000     1.052064    1.198022
               age |   .9917069   .0017098    -4.83   0.000     .9883614    .9950637
```

(a) How do unemployment claim, wage and age affect the hazard rate? **(15 points)**

(b) State which distribution has characteristics that fit in both Proportional Hazard (PH) and Accelerated Failure Time (AFT) models. **(5 points)**