AFRICAN ECONOMIC RESEARCH CONSORTIUM

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

JOINT FACILITY FOR ELECTIVES (JFE)

JUNE – SEPTEMBER 2007

ECONOMETRICS THEORY & PRACTICE I

First Semester: Final Examination

Duration: 3 Hours **Date: Wednesday, August 8, 2007**

INSTRUCTIONS:

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

You can use unprogammable calculators. Relevant formulas are embedded in the questions wherever they are necessary. Please show your derivations and mathematical steps in detail.

Question 1

1.1 Write a critique of the linear probability model (LPM) as opposed to other binary discrete choice models. **(30 points)**

1.2 Suppose, we are interested to estimate the parameters of binary discrete choice model with a dependent variable y_i. Construct the log likelihood function that needs to be maximised if p_i is the probability that $y_i = 1$. **(40 points)**

1.3 Consider a class of binary response models of the form $P(y=1|x) = G(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k) = G(\beta_0 + x\beta) = G(Z)$ and the G(.) follows a logistic distribution. How can you compute predicted probabilities for this logit model (say at the mean values of the regressors)? **(30 points)**

Question 2

2.1. Describe and show the steps of conducting the following tests;

(a). Likelihood ratio (LR) test for non-normality in probit models. **(15 points)**

(b). Hausman specification test to discriminate between fixed effects (FE) and random effects (RE) panel data models. **(15 points)**

2.2. The nested (hierarchical) model relaxes the IIA assumption of the multinomial logit model. For two-level nested model,

(a). Write down the relevant probabilities (i.e. joint, conditional and choice probabilities) and the inclusive values.

(b). Using your results under (a), derive the log likelihood function of the model and indicate how it can be estimated.

(40 points)

2.3. Students of CMAP (2006) were asked to give their assessment of the degree of difficulty of the econometrics course they are registered for. They also supplied information on their individual characteristics such as age, gender and country of origin. Their responses about the degree of difficulty of the course were coded as follows; 0=difficult; 1= manageable and 2=easy. After regressing their responses on their characteristics, we obtained the following estimated results for an ordered logit model. Note that gender is coded as 1 for females and 0 otherwise.

```
Ordered logistic regression                        Number of obs   =  33
                                                   LR chi2(5)      =  8.27

                                                Prob > chi2     =     0.1419
Log likelihood = -31.078009                     Pseudo R2       =     0.1175

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .1333478   .0774449     1.72   0.085    -.0184415     .285137
      gender |  -1.964297     1.1262    -1.74   0.081    -4.171608    .2430143
       ghana |   .8804133   1.319286     0.67   0.505    -1.705341    3.466167
      malawi |  -1.788233   1.376637    -1.30   0.194    -4.486393    .9099262
    ethiopia |  -.2106586   .8151742    -0.26   0.796    -1.808371    1.387054
-------------+----------------------------------------------------------------
       /cut1 |   1.666962   2.248936                     -2.740871    6.074795
       /cut2 |   3.092772   2.308236                     -1.431287    7.616832
------------------------------------------------------------------------------
```

(a). Interpret each of the coefficients and the threshold parameters. **(10 points)**

(b). What is the probability that a Ghanaian student is saying the course is ‘difficult’? Compare your result with the probability of a student from Malawi classifying the course as ‘difficult’? Give interpretation of your comparison.

(20 points)

Question 3

3.1. Consider a latent relationship, for observation i, of the form

$$y_i* = x_i^{'}\beta + u_i, u_i \sim N(0,\sigma^2) \qquad (2)$$

and the observability rule is such that $y_i = y_i*.1(if, y_i* > 0)$.

(a). Compute the probabilities of observing both censored and uncensored observations. **(30 points)**

(b). Derive the log-likelihood of the model. **(20 points)**

3.2.

(a). Describe the steps of generating the Heckman two-step estimator. **(20 points)**

(b). What is the inverse Mills ratio? **(15 points)**

(c). Give a real world example under which the model can be applied. **(15 points)**

Question 4

4.1. Consider the following panel data model:

$$y_{it} = \beta_0 + \gamma Z_{it} + \beta x_{it} + u_i + \varepsilon_{it}, \forall i = 1,...., n; t = 1,...., T$$

Where $u_i, and, \varepsilon_{it}$ are mutually independent unobservables; independent of Z_{it} (capital Z) (which is the measure of years of schooling completed by an individual) and other observed covariates x_{it} (e.g. age, experience and location). Note also that $\varepsilon_{it} \sim iidN(0,\sigma_\varepsilon^2)$ and you have a panel data of earnings (y_{it}) for employed individuals.

(a). Give two variables that can be captured by the unobservables u_i. **(10 points)**

(b). In the context of the human capital literature, what is the interpretation of the coefficient estimate $\hat{\gamma}$? **(10 points)**

(c). If you want to estimate a linear earnings model, you need to be concerned at least by the presence of u_i. A colleague suggest that fixed effects (FE) provides the best solution. Give a detailed description and critique of the FE estimation approach. **(30 points)**

4.2. Describe the Levin and Lin (LL-2001) panel unit root test. **(50 points)**

Question 5

5.1 Define the following duration concepts:

(a). failure function

(b). baseline hazard function

(c). shape parameter

(d). 'left censored' and 'right censored' survival times

(30 points)

5.2.

(a). Let $T \geq 0$ be the duration which has some distribution in the population and t is a particular value of T. Show that

$$\lambda(t) = \lim_{dt \downarrow 0}\left\{\frac{\Pr(t \leq T \leq t + dt \mid T \geq t}{dt}\right\} = \frac{f(t)}{S(t)}.$$

(30 points)

(b). Give real world examples for positive and negative duration dependence.
(10 points)

5.3 Suppose the Weibull model hazard rate is given by;

$$\lambda(t, x) = \alpha t^{\alpha - 1}\exp(X'\beta)$$

(a). Under what conditions do all the results of the Weibull model apply to the exponential model? **(10 points)**

(b). Let x_k is the kth covariate in the vector of characteristics X. Show how each coefficient summarises the proportionate response of the hazard to a small change in the relevant covariate. **(20 points)**