AERC
CREA

AFRICAN ECONOMIC RESEARCH CONSORTIUM

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

JOINT FACILITY FOR ELECTIVES
JULY – OCTOBER 2006

ECONOMETRICS THEORY & PRACTICE II

Second Semester: Final Examination

Duration: 3 Hours **Date: Monday, October 2, 2006**

INSTRUCTIONS:

Choose **4** out of the following **6 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 In the framework of a linear probability model(LPM), the conditional probability of an event (y) happening is given by;

$$\Pr(y=1\,|\,x)=F(x,\beta)=x'\beta$$

Introducing disturbances u, we can write the model as

$$y=x'\beta+u$$

For n observations we have;

$$y_i=x_i'\beta+u_i$$

where i is indexing individuals.

Show that the disturbances of the LPM are

(i) non-normal **(10 points)**

(ii) heteroscedastic **(20 points)**

1.2 Suppose an individual is faced with a decision whether to participate in a labour market. Show that the latent variable approach to binary choice model specification can be derived from an economic model of behaviour.

(30 points)

1.3 Consider the following binary response model;

$$P(y=1|z)=G(\beta_0+\beta_1 z_1+\beta_2 z_1^2+\beta_3\log(z_2)+\beta_4 z_3)$$

(i) Compute the marginal effects of z_1 and z_2. **(15 points)**

(ii) Give interpretations of the different components of the following logit model that predicts the probability of having a debilitating disease (i.e. *Dtfti*) for 3900 individuals. **(25 points)**

```
Logistic regression                               Number of obs   =       3900
                                                  LR chi2(10)     =      46.98
                                                  Prob > chi2     =     0.0000
Log likelihood = -2595.0551                       Pseudo R2       =     0.0090
```

Dtfti	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Old	.1485002	.0768532	1.93	0.053	-.0021293	.2991297
North	.4514692	.278616	1.62	0.105	-.094608	.9975464
South	.1040395	.0700276	1.49	0.137	-.0332121	.2412911
Female	.3960794	.0967245	4.09	0.000	.2065029	.585656
Oromo	.1630668	.076221	2.14	0.032	.0136763	.3124573
Amhara	.4112997	.1409115	2.92	0.004	.1351182	.6874812
Tigre	.2555858	.1185773	2.16	0.031	.0231786	.487993
Muslim	-.1077324	.0869757	-1.24	0.215	-.2782016	.0627368
Water	.1660216	.0947671	1.75	0.080	-.0197184	.3517617
Habit	.183462	.1958315	0.94	0.349	-.2003606	.5672847
_cons	-1.893852	.3687497	-5.14	0.000	-2.616588	-1.171115

N.B.: Marginal effects are given on the next page.

The following gives descriptions of the 10 regressors used in the above regression;

Old = a dummy variable which takes a value of 1 if the individual is aged 55 and above

North = a location dummy variable which is 1 if the individual lives in the North part of the country.

South =a location dummy variable which is 1 if the individual lives in the South part of the country.

N.B. Central is the omitted location dummy.

Female = a dummy which is 1 if the individual is female

Oromo = a dummy which is 1 if the individual is from the Oromo ethnic group

Amhara = 1 if he/she is from the Amhara ethnic group

Tigre = 1 if he/she is from the Tigre ethnic group

N.B. Gurage is the omitted ethnic group dummy.

Muslim = 1 if he/she is a Muslim

Water = 1 if the individual's access to clean water is very limited

Habit = 1 if the individual has a bad body cleaning habit(e.g infrequent hand/face washing)

In the following table, variables x1 to x10 represent the above defined variables in a respective order. Therefore, x1=old, x2=North,...etc.

```
Marginal effects after logit
      y  = Pr(tfti) (predict)
         =  .39462817
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
x1*	.0356761	.01855	1.92	0.055	-.00069	.072042	.300769
x2*	.10156	.05807	1.75	0.080	-.012254	.215374	.983077
x3*	.0248553	.01673	1.49	0.137	-.007927	.057638	.49
x4*	.0914726	.02143	4.27	0.000	.049475	.13347	.827436
x5*	.0386858	.01794	2.16	0.031	.003518	.073854	.682821
x6*	.1003044	.03479	2.88	0.004	.032109	.1685	.174359
x7*	.0601856	.02746	2.19	0.028	.006362	.114009	.731026
x8*	-.0258979	.02103	-1.23	0.218	-.067112	.015316	.794615
x9*	.0391699	.02205	1.78	0.076	-.004049	.082389	.82359
x10*	.0429415	.04479	0.96	0.338	-.044854	.130737	.968718

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Question 2

2.1

(i) Describe the mlogit (multinomial logit), clogit(conditional logit) and nlogit(nested logit) models motivating your discussion using the latent variable approach. Provide real world examples under which each of the models can be applied. **(30 points)**

(ii) What is IIA (Independence of Irrelevant Alternatives)? Show the steps of testing for IIA? **(20 points)**

2.2 Consider a sample of data $\{y_i, x_i\}$ of size n drawn independently from some population, where the dependent variable y_i has M possible outcomes $y_i = 1,...,M$ with a natural ordering. The observed values are assumed to derive some unobservable latent variable y_i^* where,

$$y_i^* = x'_i \beta + u_i, for, i = 1,...,n$$

for some kx1 parameter vector β and (univariate) stochastic disturbance term u_i. The M outcomes for the observed variable y_i are assumed to be related to the latent variable through the following observability criterion;

$$y_i = m, if, \alpha_{m-1} \leq y_i^* \leq \alpha_m, for, m = 1,...,M,$$

for a set of parameters α_0 to α_M, $\alpha_0 < \alpha_1 < \alpha_2 < ... < \alpha_M$, $\alpha_0 = -\infty$ and $\alpha_M = \infty$.

(i) Find the conditional probability of observing the mth category.
(15 points)

(ii) Derive the likelihood function for the ordered probit model. Provide examples of where the ordered probit model may be used.
(35 points)

Question 3

3.1 Consider the following observability rules of truncated and censored samples

For truncated samples, we have;

$$T1: y = y^*, if, y^* > c \text{; Not observed otherwise}$$

If the variable is censored, we have the following observability rules:

$$C1: y = y^*.1(y^* > c) + c.1(y^* \leq c)$$

$$C2: y = y^*.1(y^* < d) + d.1(y^* \geq d)$$

$$C3: y = y^*.1(c < y^* < d) + c.1(y^* \leq c) + d.1(y^* \geq d)$$

Interpret each rule and give examples in which they can be applicable.
(25 points)

3.2 Suppose two JFE students have similar econometric modelling problems but they are interested in investigating different research questions. Student A is interested primarily to predict determinants of health expenditure by households while student B is interested primarily to predict the determinant of household expenditure on environmental protection. Note that the sample of households that are going to be analysed are selected non-randomly. Furthermore, to gather relevant information the students have interviewed households in a given community about their willingness to pay a certain amount of money for health insurance as well as environmental protection. What is the relevant econometric modelling strategy for the students? Show in detail the steps involved in obtaining the parameters of interest. **(50 points)**

3.3 Given the model of the following form, $y_i = x_i'\beta + u_i, u_i \sim N(0,\sigma^2)$; show that $E(y_i \mid x_i, y_i > 0) = x_i'\beta + \sigma.\lambda(x_i'\beta / \sigma)$. **(10 points)**

3.4 Let us consider the following simultaneous equation system;

$$y_1^* = \alpha_1 y_2^* + x_1'\beta_1 + \varepsilon_1$$

$$y_2^* = \alpha_2 y_1^* + x_2'\beta_2 + \varepsilon_2$$

It is obvious that we use 2SLS technique to estimate the system if both variables on the left-hand side are completely observed. However, show how we can estimate the parameters of the system if y_1 is completely observed and y_2 is censored. **(15 points)**

Question 4

4.1 Define the following duration concepts:

(i) hazard function
(ii) survivor function
(iii) positive and negative duration dependence

(6 points)

4.2

(i) Show that $\lambda(t) = -\dfrac{d \log S(t)}{dt}$, where t is an actual realisation of a given duration T (T>t). **(10 points)**

(ii) The integrated hazard, $\Lambda(t)$, is precisely the negative of the log survival function. **(4 points)**

(iii) Given a weibull and log-logistic distributions for the shape of the hazard function, find the expression for F(t), S(t), f(t) and $\Lambda(t)$. The weibull distribution's hazard function is given by $\lambda(t) = f(t)/S(t) = \gamma\alpha t^{\alpha-1}$ while the log-logistic hazard function is given by $\lambda(t) = \dfrac{\gamma\alpha t^{\alpha-1}}{1+\gamma t^{\alpha}}$.

(15 points)

4.3 Let $T \geq 0$ be the duration of a state (e.g. unemployment, imprisonment, illness…etc) which has some distribution in the population and t is a particular value of T. Assume that T is continuous and has a differentiable cdf.

(i) Generate a formula for an average probability of leaving a given state (exit) per unit of time within the short interval.

(ii) Show the expression to compute the instantaneous rate of leaving per unit of time conditional on survival to time t.

(iii) If T is the length of time unemployed, how do you interpret $\lambda(20)$?

(15 points)

4.4 An especially important class of duration models with time-invariant regressors consists of proportional hazard models. Suppose a proportional hazard function depending on a vector of explanatory variables x with unknown coefficients is given by;

$$\lambda(t;x) = k(x) * \lambda_0(t)$$

where $k(.) > 0$ is a nonnegative function of x and and $\lambda_0(t) > 0$ is called the baseline hazard. Suppose k(.) is parameterised as $k(x) = \exp(x\beta)$ where β is a vector of parameters.

(i) Show that the proportional effect of x on the conditional probability of ending a spell does not depend on duration (i.e. t) and this effect is constant.

(10 points)

(ii) Suppose the completed durations are ordered as follows $t_1 < t_2 < ... < t_n$ and $k(.) = \phi(x, \beta)$. Derive the log-likelihood for the partial likelihood estimator suggested by Cox. Briefly discuss the advantages and shortcomings of the Cox's Proportional Hazard model.

(40 points)

Question 5

5.1 Discuss the advantages of a panel data over cross-sectional and time-series data. Highlight some of the practical (empirical) problems encountered in the process of modelling economic relationships using panel data.

(30 points)

5.2 Suppose we have the following fixed effects model with 2 time periods;

$$y_{it} = \beta_1 + \gamma_1 d2_t + \beta_2 x_{it} + a_i + u_{it}$$

where $d2_t$ is a time dummy relating to period 2; x_{it} is a time-varying explanatory variable and i can be an individual, household, firm or country identifier.

(i) Give an interpretation for a_i, and, u_{it}. **(10 points)**

(ii) What can be captured by including the time dummy? Give a real world example. **(10 points)**

(iii) Given 2 years of data, how can we estimate the β's? State key assumptions for consistent estimation of the parameters.

(50 points)

Question 6

6.1 Explain the Residual-based LM Test which is derived by McCoskey and Kao (1998) for the null of cointegration rather than the null of no cointegration in panels. **(50 points)**

6.2 What are the drawbacks of first-differencing in panel data?

(20 points)

6.3 Compare and contrast Fixed Effects (FE) and Random Effects (RE) model. State the necessary assumptions. How can one decide whether to fit an FE or RE model?

(30 points)