AFRICAN ECONOMIC RESEARCH CONSORTIUM

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

JOINT FACILITY FOR ELECTIVES (JFE) 2012

JUNE – SEPTEMBER

ECONOMETRICS THEORY AND PRACTICE II

Second Semester: Final Examination

Duration: 3 Hours **Date: Thursday, September 20, 2012**

INSTRUCTIONS:

1. This examination is divided into two sections: **Section I** and **Section II**. There is one (1) question in Section I which is compulsory, and four (4) questions in Section II.
2. You are required to attempt **THREE (3)** questions in total: **ONE (1)** question from **Section I** (**Compulsory**) and **ANY TWO** questions from **Section II**.
3. Each question carries twenty (20) marks.
4. Relevant formulae are embedded in the questions wherever they are necessary.
5. You may use an unprogrammable calculator.

Section I:

The Question in this Section is Compulsory

Question 1 [20 marks]

Suppose a researcher is interested in explaining whether people in a certain city buy health insurance or not. A sample of observations was drawn to undertake the study.

Let $y_i^* = x_i'\beta + \varepsilon_i$ i = 1, 2, …, N (1.1)

$y_i = 1$ if $y_i^* > 0$

$y_i = 0$ if $y_i^* \leq 0$

where y_i^* is the net utility to own health insurance and is not observed

$y_i = 1$ if a person has health insurance; 0, otherwise

x_i = vector of explanatory variables

The unobservable random errors, $\varepsilon_i \sim NID\,(0, 1)$.

(a) Derive the probability that a person has health insurance, i.e., P{ $y_i = 1 | x_i$ }. [Note: By symmetry P{ $\varepsilon_i > -x_i'\beta$ } = P{ $\varepsilon_i < x_i'\beta$ }.] **[3 marks]**

(b) Derive the probability that a person does not have health insurance, i.e., $P\{y_i = 0 | x_i\}$. **[1 marks]**

(c) What specific model is appropriate to estimate the model in equation (1.1)? Why? **[2 marks]**

(d) Derive the log-likelihood function (log-L) for this model. **[4 marks]**

(e) Derive the log-likelihood equations for this model, i.e., the first-order conditions for maximizing Log-L. **[2 marks]**

(f) Briefly state the steps to estimate this model using maximum likelihood estimation. State the necessary and sufficient conditions to find the optimum β. **[3 marks]**

(g) In this probability model, the interpretation of the parameter β is not easy and instead, the marginal effects are used to interpret results. Write the expression and interpretation for marginal effects. **[1 marks]**

(h) Suppose the random error terms (ε_i) are assumed, instead to have a logistic distribution where μ = 0 and $\sigma^2 = \pi^2/3$. How would you propose the model in equation (1.1) to be estimated? **[1 marks]**

(i) Suppose the researcher decided to choose the linear probability model (LPM) to estimate the model in equation (1.1). Do you agree on the choice of LPM as an appropriate estimator for equation (1.1)? Explain why or why not? **[3 marks]**

Section II:

Answer ANY TWO Questions from this Section

Question 2 [20 marks]

In the context of multinomial models, the individuals may face more than two alternatives to choose from. Each individual chooses the alternative which provides the greatest utility. Depending on the type of observations on the explanatory variables (choice attributes, z_{ij}, or individual characteristics, w_i, where index *i* refers to individual and *j* stands for choice categories), the probability of the individual choosing *j*, *j* = 1, 2, …, *J* may be expressed as:

$$P\{y_i = j\} = \frac{\exp\{z_{ij}'\gamma\}}{1 + \exp\{z_{i2}'\gamma\} + \ldots + \exp\{z_{iJ}'\gamma\}} \qquad (2.1)$$

for conditional logit, or

$$P\{y_i = j\} = \frac{\exp\{w_i'\alpha_j\}}{1 + \exp\{w_i'\alpha_2\} + \ldots + \exp\{w_i'\alpha_J\}} \qquad (2.2)$$

for multinomial logit.

(a) Explain the difference between the conditional logit and the multinomial logit models. **[4 marks]**

(b) In order to use the conditional logit and multinomial logit, what restriction is implicitly imposed on the odd-ratios in relation to different *J* alternatives? If this assumption fails, what alternative models may be estimated? Briefly describe each model. **[5 marks]**

(c) Explain the difference between the multinomial logit model and the ordered logit model. **[3 marks]**

(d) Suppose the ordered logit model is used to analyze the factors that influence food security status in Kenya based on some latent variable, $y_i^* = x_i'\beta + \varepsilon_i$, where x_i = (income, market distance, rainfall, conflict)′. The food insecurity status, y_i, is categorized and assigned values 1, 2, 3, and 4 such that:

$y_i = 1$ [extremely food insecure], if $y_i^* \le \delta_1$

$y_i = 2$ [highly food insecure], if $\delta_1 < y_i^* \le \delta_2$

$y_i = 3$ [moderately food insecure], if $\delta_2 < y_i^* \le \delta_3$

$y_i = 4$ [generally food secure], if $y_i^* > \delta_3$

(i) Derive the probabilities of observing each category 1, 2, 3, and 4, and express those probabilities in terms of the logistic cumulative distribution function, $\Lambda(.)$. **[4 marks]**

(ii) If income is increased by one unit, intuitively what would you expect to happen in the probabilities of observing a household to belong to extremely food insecure and generally food secure groups? **[2 marks]**

(iii) Conflict is a dummy which takes the value 1 if there is a presence of conflict, 0 otherwise. If conflict changed from 0 to 1, intuitively what would you expect to happen in the probabilities of observing a household to belong to extremely food insecure and generally food secure groups? **[2 marks]**

Question 3 [20 marks]

Consider the following model with given specifications and assumptions:

$$w_i^* = x_{1i}'\beta_1 + \varepsilon_{1i} \qquad \text{i} = 1, 2, \ldots, N \qquad (3.1a)$$

$$h_i^* = x_{2i}'\beta_2 + \varepsilon_{2i} \qquad (3.1b)$$

$$w_i = w_i^*, \qquad h_i = 1 \text{ if } h_i^* > 0 \qquad (3.1c)$$

$$w_i \text{ not observed}, \; h_i = 0 \text{ if } h_i^* \leq 0$$

and

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim NID\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix}\right] \qquad (3.1d)$$

where w_i^* denotes person i 's potential wage and is not observed for people who are not working.

w_i denotes person i 's actual wage

h_i^* indicates net utility

h_i indicates working and not working

(a) What type of model will you choose to appropriately match the specifications and assumptions of the model above? Derive the probability that $w_i = 0$, and $E\{w_i \mid h_i = 1\}$.

[Hint: $E\{\varepsilon_{1i} \mid h_i = 1\} = \sigma_{12} \dfrac{\phi(x_{2i}'\beta_2)}{\Phi(x_{2i}'\beta_2)}$] **[5 marks]**

(b) If $x_{1i}'\beta_1 = x_{2i}'\beta_2$ and $\varepsilon_{1i} = \varepsilon_{2i}$, then what type of model is appropriate for estimation? Derive the probability that $w_i = 0$, and $E\{w_i \mid w_i > 0\}$.

[Hint: $E\{\varepsilon_{1i} \mid \varepsilon_{1i} > -x_{1i}'\beta\} = \sigma_1 \dfrac{\phi(x_{1i}'\beta_1/\sigma_1)}{\Phi(x_{1i}'\beta_1/\sigma_1)}$] **[5 marks]**

(c) Explain the differences among the censored regression, truncated regression, and the sample selection models. **[3 marks]**

(d) Under the model (3.1) specification, the conditional expected wage, given that the person is working, is given by:

$$E\{w_i \mid h_i = 1\} = x_{1i}'\beta_1 + \sigma_{12}\frac{\phi(x_{2i}'\beta_2)}{\Phi(x_{2i}'\beta_2)} \qquad (3.2)$$

Equation (3.2) may be estimated using OLS under a certain condition. In which condition will OLS yield a consistent estimate of β? **[3 marks]**

(e) Equation (3.2) may also be expressed as:

$$w_i = x_{1i}'\beta_1 + \sigma_{12}\lambda_i + \eta_i \qquad (3.3)$$

where $\lambda_i = \dfrac{\phi(x_{2i}'\beta_2)}{\Phi(x_{2i}'\beta_2)}$

$$\eta_i = \varepsilon_{1i} - E\{\varepsilon_{1i} \mid x_i, h_i = 1\}$$

Equation (3.3) may be estimated using the Heckman's 2-step estimation procedure. Discuss briefly the steps involve in this procedure. **[4 marks]**

Question 4 [20 marks]

Panel data refers to the pooling of cross-section observations over several periods. Suppose there are observations on *N* individuals, *i* = 1, 2, ..., *N*, taken over *T* periods of time, *t* = 1, 2, ..., *T*. Consider the following panel data model where the dependent variable, y_{it} may be described as a linear function of the individual's previous realizations, and of explanatory variables, x_{it} based on some fundamental economic theory.

If the relationship is static, the econometric expression maybe written as:

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it} \qquad (4.1)$$

where: $u_{it} \sim IID(0, \sigma_u^2)$, and x_{it} is uncorrelated with u_{it}.

(a) What are the conventional assumptions on the behavior of the individual-specific effects, α_i to be considered before estimating the model in equation (4.1)? **[2 marks]**

(b) Depending on the assumptions on the behavior of α_i, what estimators would you suggest to consistently and efficiently estimate the model in equation (4.1)? **[4 marks]**

(c) One of the many advantages of using panel data is that it is able to address the problem of omitted variable bias. Show how a fixed-effect model is able to rectify this problem. [You may use an example to support your argument.] **[5 marks]**

(d) Suppose y_{it} is the production output in 35 African countries from 2001 to 2004, and the corresponding $x_{it} = (x_{1it}, x_{2it})'$ are production inputs, respectively, capital and labor. All variables are expressed in logs. The coefficients from the estimated model with constant returns to scale, and some test results are presented below.

Table 1: Panel data estimates of the growth equation

Dependent variable: log of production	*FE* Coefficients (*t*-values)	*RE* Coefficients (*t*-values)
Constant	2.123*** (9.66)	1.820*** (5.73)
Elasticity of output with respect to capital	0.371*** (8.25)	0.447*** (10.67)
Elasticity of output with respect to labor	0.629*** (13.98)	0.553*** (6.29)
Adjusted R-squared	0.99	0.73
No. of cross-sections	35	35
Total no. of observations	140	140
Hausman test: $\chi^2(2) = 18.70$***		
White test heteroskedasticity: $\chi^2(5) = 4.87$		

Figures in parenthesis are *t*-values. *** Significance at 1%, ** Significance at 5 %

(i) What conclusion can you draw from the computed value of the Hausman test? State the null and alternative hypotheses. **[3 marks]**

(ii) What conclusion can you draw from the heteroskedasticity test? **[2 marks]**

(iii) Interpret the coefficients of the model of your choice based on the tests (i) and (ii) above. **[4 marks]**

Question 5 [20 marks]

Consider the linear dynamic panel data model with exogenous variables and a lagged dependent variable given by:

$$y_{it} = x_{it}'\beta + \gamma y_{i,t-1} + \alpha_i + u_{it} \qquad i = 1, 2, \ldots, N \qquad t = 1, 2, \ldots, T \qquad (5.1)$$

where $u_{it} \sim IID(0, \sigma_u^2)$.

(a) If the individual-specific effect is constant over time, would you estimate the model in equation (5.1) using the fixed effect estimator? Why or why not? **[3 marks]**

(b) If the fixed-effects estimator is not recommended to estimate (5.1), what alternative estimators would you use? **[2 marks]**

(c) Consider the dynamic panel data model without the exogenous variables given by:

$$y_{it} = \gamma y_{i,t-1} + \alpha_i + u_{it}, \qquad |\gamma| < 1 \qquad (5.2)$$

Arellano-Hsiao suggested two estimators to find consistent estimates. These are given below:

$$\hat{\gamma}_{IV}^{(1)} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i,t-2}(y_{it} - y_{i,t-1})}{\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i,t-2}(y_{i,t-1} - y_{i,t-2})} \qquad (5.3)$$

and

$$\hat{\gamma}_{IV}^{(2)} = \frac{\sum_{i=1}^{N}\sum_{t=3}^{T} (y_{i,t-2} - y_{i,t-3})(y_{it} - y_{i,t-1})}{\sum_{i=1}^{N}\sum_{t=3}^{T} (y_{i,t-2} - y_{i,t-3})(y_{i,t-1} - y_{i,t-2})} \qquad (5.4)$$

What are the instrumental variables used in (5.3) and (5.4)? What are the necessary conditions for consistency for each estimator? **[6 marks]**

(d) If we use the matrix of instruments given in Z_i below, in the estimation of the dynamic panel, what are the problems that one is likely to encounter?

$$Z_i = \begin{pmatrix} [y_{i0}] & 0 & \cdots & 0 \\ 0 & [y_{i0}, y_{i1}] & & 0 \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & [y_{i0}, \ldots, y_{i,T-2}] \end{pmatrix}$$

[4 marks]

(e) If we are dealing with the panel time series, why is it important to do panel unit root tests before model estimation? **[3 marks]**

(f) Explain the difference between the Levin and Lin panel unit root test, and the Im, Pesaran and Shin test. **[2 marks]**