AFRICAN ECONOMIC RESEARCH CONSORTIUM

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

JOINT FACILITY FOR ELECTIVES (JFE) 2016
JUNE – SEPTEMBER

ECONOMETRICS THEORY AND PRACTICE II

Second Semester: Final Examination

Duration: 3 Hours **Date: Thursday, September 22, 2016**

INSTRUCTIONS:

1. This examination is divided into two parts: Section A and Section B.
2. You are required to answer **ANY TWO QUESTIONS** in **EACH SECTION**.
3. All questions carry 15 marks each.

Section A:

Answer ANY TWO Questions in this Section

Question 1

A researcher considers the panel data regression model

$$y_{it} = \phi + X_{it}'\beta + \alpha_i + v_{it}, \quad i = 1,\ldots,N, \quad t = 1,\ldots,T \tag{1}$$

where i and t denote the cross-section and time-series dimensions respectively, whereas α_i represents the unobservable individual-specific effect and v_{it} the remainder disturbance. ϕ is a scalar, β a $K \times 1$ vector of coefficients and X_{it} the related set of explanatory variables.

(a) Is the ordinary least squares (OLS) an appropriate method for the estimation of β? Why or why not? If not, what are the alternative estimation approaches? **[3 marks]**

(b) After providing some advantages and drawbacks of the two major alternatives to OLS in running the regression (1), explain how one can discriminate among these two models. **[6 marks]**

(c) The researcher decides to apply the random effect approach to model (1). He therefore gathers the individual effect and the remainder error term in a single error u_{it} as follows:

$$u_{it} = \alpha_i + v_{it} \tag{2}$$

(i) Show that, in matrix form, $u = Z_\alpha \alpha + v$ where the matrix Z_α would be defined. **[3 marks]**

(ii) After setting the appropriate assumptions regarding the error components α_i and v_{it}, determine the spectral decomposition of the variance-covariance matrix of the error vector u. All the matrices involved will have to be defined accordingly. **[3 marks]**

Question 2

An insecticide is spread in quantity γ in a closed box containing a sample of N observed insects. Let $y_i = 1$ if insect i is found dead ten minutes later and $y_i = 0$ if not. Let y_i^* be an unobserved tolerance of insect i which depends on k characteristics of the insect (weight, age, gender, etc.) gathered in a vector x_i without a constant term:

$$\forall i = 1,\dots,N,\; y_i = \begin{cases} 1 \text{ if } y_i^* < \gamma \\ 0 \text{ if } y_i^* \geq \gamma \end{cases} \text{ with } y_i^* = \beta' x_i + \varepsilon_i \qquad (1)$$

where $\varepsilon_i \sim iid\left(0, \sigma_\varepsilon^2\right)$ is the error term which is such that its standardized version has a cumulative density function $F(\cdot)$.

(a) Express the probability that the ith insect dies at the end of the experiment as a function of the explanatory factors x_i, the function $F(\cdot)$, and the parameters β, γ, and σ_ε. **[2 marks]**

(b) What are the usual models, including the linear one, suggested in the econometric literature to handle this type of problem? Briefly discuss the advantages and the drawbacks of each of them. **[6 marks]**

(c) Consider a model in which the function $F(\cdot)$ is assumed to be the cumulative density function of the logistic distribution: $F(x) = \Lambda(x) = \dfrac{e^x}{1+e^x} \quad \forall x \in \mathbb{R}$.

(i) Write down the log likelihood function of the model and derive the first order conditions. **[6 marks]**

(ii) Which comment could you make from the value of the coefficient of the age for instance? **[1 mark]**

Question 3:

Suppose we use an index formulation for a discrete choice model but it is felt that the latent variable is strictly positive. This is accommodated by assuming that the latent variable y^* has an exponential density with parameter γ. The density $f(y^*)$ is thereby such that $f(y^*) = \gamma^{-1}\exp(-y^*/\gamma)$, with $\gamma = \exp(x'\beta)$. Moreover, we observe that $y = 1$ if $y^* > z'\alpha$ and $y = 0$ if $y^* \leq z'\alpha$, where *z* is another set of covariates.

(a) Give the log-likelihood function for the observed data $\{y_i; x_i; z_i\}_{i=1,\ldots,n}$. **[5 marks]**

(b) What is the effect of a one-unit change in x_{ij} on $\Pr[y_i = 1]$, *i* denoting the individual and *j* an explanatory variable? **[5 marks]**

(c) Suppose that $y = 1$ if $y^* > \exp(z'\alpha)$ and $x = z$. Do you see any problem in identifying α and/or β? Explain your answer. **[5 marks]**

Section B:

Answer ANY TWO Questions in this Section

Question 4

Consider the Weibull distribution and its cumulative density function given by: $F(t) = 1 - \exp\left(-(\omega t)^{\gamma}\right)$ where ω and γ are non-negative parameters.

(a) Derive the related density function. **[3 marks]**

(b) Derive the hazard function. **[3 marks]**

(c) Returning to the Weibull distribution, indicate the behavior of this distribution in the following cases:

(i) $\gamma = 1$ **[3 marks]**

(ii) $\gamma > 1$ **[3 marks]**

(iii) $\gamma < 1$ **[3 marks]**

Question 5

A discrete choice model is estimated in order to investigate the determinants of the binary variable "to have at least a child" denoted by ENF, taking on the values 0 (no child) or 1 (at least one child). The highest degree or diploma of the head of the household ($DIPL = 1$ if the individual has a degree below the A level, $DIPL = 2$ if the highest degree is equivalent to the A level, and $DIPL = 3$ if he has more than the A level), the individual age (AGE) and the square of that age $\left(AGE2 = \frac{AGE^2}{100} \right)$ are the regressors to be considered. The output of the estimation with STATA is provided, with the "vce" command presenting the variance-covariance matrix of the parameters' estimates. The sample contains adults of ages between 20 to 60. The variables $\_Idipl\_1$, $\_Idipl\_2$, and $\_Idipl\_3$ are dummy variables derived from *DIPL*.

```
. xi: probit enf i.dipl age age2
i.dipl            _Idipl_1-3          (naturally coded; _Idipl_1 omitted)

Iteration 0:   log likelihood = -18636.845
Iteration 1:   log likelihood = -17371.858
Iteration 2:   log likelihood = -17369.568
Iteration 3:   log likelihood = -17369.568

Probit estimates                                  Number of obs   =      28922
                                                  LR chi2(4)      =    2534.55
                                                  Prob > chi2     =     0.0000
Log likelihood = -17369.568                       Pseudo R2       =     0.0680

------------------------------------------------------------------------------
         enf |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _Idipl_2 |  -.1324138   .0236364    -5.60    0.000    -.1787404   -.0860873
    _Idipl_3 |  -.2885005    .018667   -15.46    0.000     -.325087   -.2519139
         age |   .2996346    .006378    46.98    0.000     .2871339    .3121352
        age2 |  -.3834785   .0080627   -47.56    0.000    -.3992812   -.3676758
       _cons |  -4.966942   .1199125   -41.42    0.000    -5.201966   -4.731918
------------------------------------------------------------------------------

. vce

             | _Idipl_2 _Idipl_3      age     age2    _cons
-------------+---------------------------------------------
    _Idipl_2 |  .000559
    _Idipl_3 |  .000104  .000348
         age |  5.3e-06 -4.8e-06  .000041
        age2 | -3.7e-06  7.8e-06 -.000051  .000065
       _cons | -.000248 -.000041 -.000755  .000932  .014379
```

(a) Which model has been run? How could you obtain the regression coefficients for the other famous alternative model? Explain the factor used by Amemiya for the approximation. **[3 marks]**

(b) Would you say that the diploma has a negative or positive impact on fecundity? **[2 marks]**

(c) Test the hypothesis that the coefficients of the variables $\_Idipl\_2$ and $\_Idipl\_3$ are significantly the same? **[4 marks]**

(d) We now include a variable NENF taking on values 0, 1, 2, 3 or 4, if the number of children in the household is 0, 1, 2, 3, or at least 4, respectively. We then estimate an ordered probit model with unknown cut-offs s_j for $j = 0, 1, \ldots, 5$:

$NENF_i = j \in \{0,1,2,3,4\}$ if $s_j < x_i'\beta + u_i \leq s_{j+1}$ with $u_i \sim N(0, \sigma^2)$. We assume that $s_0 = 0$ and $s_5 = +\infty$.

The estimation output obtained with Stata is provided below (the parameters $\_cut1$, ..., $\_cut4$ denote the thresholds or cut-offs s_1, ..., s_4 of the model).

```
. xi: oprobit nenf i.dipl age age2
i.dipl            _Idipl_1-3          (naturally coded; _Idipl_1 omitted)

Iteration 0:   log likelihood = -40356.058
Iteration 1:   log likelihood = -39014.837
Iteration 2:   log likelihood = -39013.178

Ordered probit estimates                          Number of obs   =      28922
                                                  LR chi2(4)      =    2685.76
                                                  Prob > chi2     =     0.0000
Log likelihood = -39013.178                       Pseudo R2       =     0.0333

------------------------------------------------------------------------------
        nenf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _Idipl_2 |  -.1145041   .0194437    -5.89   0.000    -.1526129   -.0763952
    _Idipl_3 |  -.2033115   .0155256   -13.10   0.000    -.2337412   -.1728817
         age |     .26945   .0054888    49.09   0.000     .2586921     .280208
        age2 |  -.3453101    .006959   -49.62   0.000    -.3589496   -.3316706
-------------+----------------------------------------------------------------
       _cut1 |   4.436445   .1032947          (Ancillary parameters)
       _cut2 |   5.178941   .1041345
       _cut3 |   6.082835   .1050567
       _cut4 |    6.83505   .1058442
------------------------------------------------------------------------------
```

```
. vce

             | _Idipl_2 _Idipl_3      age     age2    _cut1    _cut2    _cut3
-------------+---------------------------------------------------------------
    _Idipl_2 |  .000378
    _Idipl_3 |  .000066  .000241
         age |  3.9e-06 -1.9e-06   .00003
        age2 | -3.1e-06  3.3e-06 -.000038  .000048
       _cut1 |  .000168  .000047  .000559 -.000692   .01067
       _cut2 |  .000167  .000043  .000564 -.000698  .010729  .010844
       _cut3 |  .000165  .000041  .000569 -.000704  .010794  .010898  .011037
       _cut4 |  .000162  .000039  .000569 -.000705  .010803  .010903  .011025

             |    _cut4
-------------+---------
       _cut4 |  .011203
```

Show that the above simple probit regression of questions (2) and (3) is a special case of the ordered probit model by computing $\Pr\left(ENF_i = 1 | x_i\right)$ as a function of the threshold s_1. Deduce that the intercept of the simple probit model depends on the opposite of s_1. **[3 marks]**

(e) Compute the estimate of the probability to have 3 children in a household whose head has a diploma below the A level and is 35 years old? <u>*Hint*</u>*: A table of the standard normal distributed is provided in Appendix 1.* **[3 marks]**

Question 6

Let x^* be $N\left(\mu,\sigma^2\right)$, $\phi(\cdot)$ and $\Phi(\cdot)$ the probability and cumulative density functions of the $N(0,1)$ distribution, respectively. In addition, the variable of interest x is defined as follows:

$$x = \begin{cases} x^* \text{ if } x^* > c \\ c \text{ if } x^* \le c \end{cases} \text{ for some constant } c.$$

(a) Verify the expression $E(x) = c\Phi\left(c^*\right) + \left(1-\Phi\left(c^*\right)\right)\left[\mu + \sigma\dfrac{\phi\left(c^*\right)}{1-\Phi\left(c^*\right)}\right]$ where c^* is the standardized value of c. **[6 marks]**

(b) Show that the variance of x is given by

$$\operatorname{Var}(x) = \sigma^2\left[1-\Phi\left(c^*\right)\right]\left[1-\delta\left(c^*\right)+\left(c^* - \frac{\phi\left(c^*\right)}{1-\Phi\left(c^*\right)}\right)^2 \Phi\left(c^*\right)\right]$$

where $\delta\left(c^*\right) = \dfrac{\phi\left(c^*\right)}{1-\Phi\left(c^*\right)}\left[\dfrac{\phi\left(c^*\right)}{1-\Phi\left(c^*\right)} - c^*\right]$ for $x > c$ or $\dfrac{-\phi\left(c^*\right)}{\Phi\left(c^*\right)}\left[\dfrac{-\phi\left(c^*\right)}{\Phi\left(c^*\right)} - c^*\right]$ for $x < c$.

[9 marks]

<u>Hints:</u> *Use the fact that* Variance $= E$(conditional variance) + Var(conditional mean) *and the formulas for conditional and unconditional means of a truncated normal random variable: The truncated density is given by*

$$f\left(x | x > c\right) = \frac{f(x)}{\Pr(x > c)} = \frac{1}{\sigma}\phi\left(\left[x-\mu\right]/\sigma\right) \Big/ \left[1-\Phi\left(\frac{c-\mu}{\sigma}\right)\right] \textit{ for } c < x < \infty, \textit{ or}$$

$f(x|x<c) = \frac{f(x)}{\Pr(x<c)} = \frac{1}{\sigma}\phi([x-\mu]/\sigma) \Big/ \Phi\left(\frac{c-\mu}{\sigma}\right)$ *for* $-\infty < x < c$ *when the truncation is above.*

The conditional means are $E(x|x>c) = \mu + \sigma\frac{\phi(c^*)}{1-\Phi(c^*)}$ *and* $E(x|x<c) = \mu - \frac{\phi(c^*)}{\Phi(c^*)}$ *regarding the direction of the truncation. The conditional variance is* $\sigma^2\left[1-\delta(c^*)\right]$.

Appendix 1: Table of the CDF of the N(0,1) distribution

z	0	1	2	3	4	5	6	7	8	9
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Examples: If $Z \sim$ Normal(0,1) then $P(Z \leq -1.32) = .0934$ and $P(Z \leq 1.84) = .9671$.
Source: This table was generated using the Stata® function normd.