AFRICAN ECONOMIC RESEARCH CONSORTIUM

Collaborative PhD Programme in Economics

JOINT FACILITY FOR ELECTIVES

JULY 1 - OCTOBER 31, 2008

ECON 661: ECONOMETRICS

First Semester: Final Examination

Time: 09:00 AM – 12:00 Noon **Tuesday, August 19, 2008**

INSTRUCTIONS:

1. The exam will be completed in 3 hours (180 minutes).
2. All questions in the exam are compulsory questions and allocation of points reflects the time you should spend in each question and sub-questions within.
3. You are only allowed to have a simple calculator, writing tools and a watch. No other material is allowed during the exam.
4. You are only allowed to write your exam with permanent ink.

Question 1 **Asymptotic Theory** **[25 Points]**

This exam is supplemented with a document that provides the basic tools required to establish the asymptotic distribution of estimated parameters that often rely on finite samples. You may like to use the provided information to answer the following questions:

(a) Explain – one by one and with as few words as possible – the consequence of the breakdown of each of the assumptions A1 to A6. **[2 points]**

(b) Econometricians are often concerned with the study of the asymptotic distribution of some measurable distance $(\hat{\theta}_n - \theta_0)$ where $\hat{\theta}_n$ is based on some finite sample. Justify the importance of studying such asymptotic distribution and furthermore motivate the reason why the distance $(\hat{\theta}_n - \theta_0)$ needs to be normalized by $\sqrt{n}$ when we search for its asymptotic distribution. Be brief and to the point when you answer this question.
[2 points]

(c) Let's consider a Linear Regression Model $y_i = x_i'\beta + u_i$ that may also be expressed in Matrix form as $Y = X\beta + U$ where $x \in X \subseteq \mathbb{R}^k$. It is assume that the information set $\{z_i\} = \{y_i, x_i\}$ for $i = 1,...,n$ is iid with finite 1st and 2nd moments. The sample of size n is assumed to be a finite and representative of the underlying population. You may use sample analogue or matrix format to answer the following questions:

(i) Explicitly use Assumptions A1-A5 to show that the parameter vector resulting from least square estimation ($\hat{\beta}_{ols}$) is unbiassed, and provide an expression that explains the estimated parameter net from its true population value β_o. What is the dimension of such parameter? **[2 points]**

(ii) Given that the estimated parameter is based on some finite sample, what do we imply by suggesting that such parameter is 'unbiassed'? Be brief and to the point in your answer. **[2 points]**

(iii) Explain why Lindeberg-Levy is the appropriate CLT to be applied in this case and, together with other appropriate asymptotic tools, use such CLT to establish the distributional properties of the sum $\sum_{i=1}^{n} x_i u_i$ that should appear in the expression you have derived in (i). **[3 points]**

(iv) Use the outcome in (iii) together with the outcome in (i) and Crammer's Theorem to finally derive the asymptotic distribution of the measurable distance $\sqrt{n}(\beta_{ols} - \beta_o)$. **[3 points]**

(v) It is well known that the estimated variance in the regression $\hat{\sigma}_2$ can be explained as $(n-k)\frac{\hat{\sigma}^2}{\sigma^2} = \left(\frac{u}{\sigma}\right)' M \left(\frac{u}{\sigma}\right)$ where $M = [I - X(X'X)X']$. Show that M is symmetric and idempotent and use the given information for $\hat{\sigma}_2$ together with the expression you have derived in (i) to demonstrate that the parameters $\hat{\beta}_{ols}$ and $\hat{\sigma}_2$ are independent. **[4 points]**

(vi) Let $k = 1$ so that now the model is simply given by $y_i = \alpha + x_i\beta_1 + u_i$; use your answer to (iv) to simply write down the distribution of $\frac{\sqrt{n}(\hat{\beta}_1 - \beta_1)}{\sigma\left(\sum x'x\right)^{-1/2}}$. **[1 point]**

(vii) Clearly, σ is unknown so that we would need to replace it by the estimated parameter such that $\frac{\sqrt{n}(\hat{\beta}_1 - \beta_1)}{\hat{\sigma}\left(\sum x'x\right)^{-1/2}}$ would be our consideration with an identical distribution as that derived in (vi). It is also well known that $\frac{\sigma\left(\sum x'x\right)^{-1/2}}{\hat{\sigma}\left(\sum x'x\right)^{-1/2}} \xrightarrow{p} 1$. Use Fact 1 in the supplementary sheet together with information you have just read to show that the $Student's - t$ is a pivotal statistics. **[6 points]**

Question 2 Maximum Likelihood Estimation [25 Points]

(a) What do we mean when we say that $\hat{\theta}_{mle}$ is the most likely admissible vector of parameters for some fixed information set $\{z_i\},\ i=1,...,n$? Use your explanation to justify the principle behind Maximum Likelihood estimation. Be brief and concise in your answer.

[1 point]

Let $\{z_i\}=\{y_i,x_i\},\ i=1,...,n$ be a sequence of iid random draws from which we have a sample of size n that represents the underlying population. The outcome y_i is binary such that $y_i=1$ if $(x_i\beta+u_i)>0$ and zero otherwise.

(b) In the case that $u_i \sim N(0,\sigma^2)$ show that $P(y|x)$ is a proper Joint density function (Hint: you have to show that the mass under the density adds up to one). **[1 point]**

(c) In the case that $u_i \sim \Lambda(0,s^2)$ (i.e., the Logistic distribution), $P(y=0|x)=1/(1+\exp(x\beta))$. Use it to derive $P(y=1|x)$ given that $P(y|x)$ is a proper density function. **[1 point]**

(d) Write down the likelihood function $L(\beta;z_i)$ and its log-transformation given $u_i \sim N(0,\sigma^2)$. Do the same in the case that $u_i \sim \Lambda(0,s^2)$. Note that in both cases you must not derive these functions but straightforwardly write them down! **[2 points]**

(e) Derive the two Score functions $q(\beta)$ from the two Log-Likelihood functions you have derived in (d). **[3 points]**

(f) Show that $Eq(\beta)=0$ **only** for the case when the score is derived from the log-likelihood given $u_i \sim N(0,\sigma^2)$. **[3 points]**

(g) Derive the Hessian (i.e., $H(\beta)=\dfrac{\partial^2 q(\beta)}{\partial\beta\partial\beta'}$) **only** for the case of the log-likelihood function derived using the logistic distribution. You may like to use the appropriate Score function derived in (e) as way of finding $H(\beta)$. **[4 points]**

(j) Assume **only** the interpretation $u_i \sim \Lambda(0,s^2)$. Use both the Score and the Hessian functions to set up the iteration procedure according to the Newton-Raphson method. Note that you must not derive but straightforwardly set up the actual iteration procedure using your results from part (e) and (g) given that we assume $u_i \sim \Lambda(0,s^2)$. **[2 points]**

(k) Why must we use iteration techniques to obtain estimates for the parameter vector β in this case? Be brief and concise in your answer. Briefly write down the steps behind such iteration technique. **[2 points]**

(k) The supplementary sheet provides the definitions and distributional properties of the Wald Test, the Lagrange Multiplier test and the Likelihood Ratio Test. Use a graphic interpretation in order to compare the methodology behind these three tests. **[3 points]**

(j) Justify why it must be the case that under the alternative (i.e., assuming that H_0 is false) $2[\log Q(\hat{\theta}_{H1}) - \log Q(\hat{\theta}_{H0})] \rightarrow \infty$. **[3 points]**

Question 3 Panel Data Strategies [25 Points]

Let $y_{it} = x'_{it}\beta + f_i + u_{it}$ where $i = 1,....,N$ for any given time period $t = 1,...,T$ thus assuming a balance panel as way of explaining the Panel Structure of the data. Answer the following questions:

(a) Show, using a graphical interpretation, that our estimates of β might be biased in a variety of ways if we assume $f_i = f$ when in fact f_i is indeed heterogeneous among individuals with ith characteristics. **[3 points]**

(b) Why must we assume 'exogeneity' (either weak or strong) between u_{it} and x_{it}? **[1 point]**

(c) What assumptions must you make between f_i and x_{it} in the event that you decide to estimate a Random Effects Model? Provide reasons for your answer. **[2 points]**

(d) What assumptions must you make between f_i and x_{it} in the event that you decide to estimate a Fixed Effects Model? Provide reasons for your answer. **[2 points]**

(e) Assume a Random Effects specification of the given model and define $v_{it} = f_i + u_{it}$ for simplicity. Furthermore assume that A3 applies to u_{it} while $E(u_{it} | f_i) = 0$ is also assumed. Derive the expression for $E(v_{it}v_{it}')$. What choice of modelling strategy would be justified when the variance covariance matrix is of such form? **[3 points]**

(f) Assume the settings for a Fixed Effect model, that is, assume that $E(u_{it} | x_{it}) = 0$ while $E(f_i | x_{it}) = g(M_i)$ for some M_i information set. Explain why it must be the case that estimating the model allowing for simple OLS or the use of a GLS would imply inconsistent estimates for the parameter vector β. **[3 points]**

(g) With the obvious notation, let $Q = \left(I_T - \frac{1}{T}ee'\right)$ be the mean-transformation matrix. Apply this transformation to the model $y_{it} = x'_{it}\beta + f_i + u_{it}$ and derive the expression for the Within Groups estimate $\hat{\beta}_{WG}$. As you derive it you must show how such model transformation eliminates the problems implied by unobserved fixed effects. **[5 points]**

(h) Assuming a RE specification it can be shown that $Var(\beta_{GLS}) = \sigma_u^2\left[\sum x_i[Q' + \psi P]x_i\right]^{-1}$ where $P = (1/T)ee'$ is an idempotent and symmetric matrix and ψ is some positive number. Likewise assuming a FE model it can be shown that $Var(\beta_{WG}) = \sigma_u^2\left[\sum x_i' Q x_i\right]^{-1}$. Show that estimation allowing for RE assumptions results in an estimated parameter vector that is as least as efficient as that implied by Within Groups estimation. **[6 points]**

Question 4 Non-parametric Programme Evaluation Econometrics [25 Points]

Assume that in a given economy a policy treatment D is put into effect such that $D_i = 1$ indicates that unit $i \in n$ has been treated while $D_i = 0$ indicates that the unit has not been treated by the policy. The sample of size n is representative of the population and can be express as $n = n_1 + n_0$ with obvious notation. At some point after the treatment has taken place we observe the realized outcome of interest Y. Whether the person is treated or not, it is a fact that for any given unit in the population we can define two potential outcomes, namely, Y_i^1 that explains the outcome we would observe in the event that the ith individual is treated and Y_i^0 that would result in the event that the ith unit is not treated by the policy. A set of covariates X explains the characteristics of the underlying population.

(a) A classic econometrician may postulate specifications $Y^1 = X\beta_1 + U_1$ and $Y^0 = X\beta_0 + U_0$ while he also assumes that once the outcome is realized, $Y = Y^1 D + (1-D)Y^0$. Use a classic econometric set up to show that the result from running a linear regression of Y on X and D is to estimate the conditional effects of the policy impact as $\alpha = E[Y_1 - Y_0 \mid X]$, as opposed to estimating the policy impact $E[Y_1 - Y_0 \mid X, D=1]$. **[4 points]**

(b) Explain one reason why the estimated parameter $\hat{\alpha}_{ols}$ you have just derived in (a) may be bias and inconsistent. (Hint: you may like to use part of A3 to examine your derivations in (a) when answering this question). **[2 points]**

(c) In general, reason (using words) why the policy value $E[Y_1 - Y_0 \mid D=1]$ may be a more appropriate measure of the policy impact than the alternative measure $E[Y_1 - Y_0]$. **[2 points]**

(d) Analytically show the identification problem associated with the policy value $ATET = E[Y_1 - Y_0 \mid D=1]$. **[2 points]**

(e) Look at the CIA in the supplementary document and explain **the meaning** and **importance** of the assumptions *(i)* $Y^1, Y^0 \perp D \mid X$, *(ii)* $X : exogenous$ and *(iii)* $0 < p(x) < 1$ at establishing identification of the parameter ATET. Be brief and to the point. **[3 points]**

(f) Explain the importance of the propensity score in the matching process and provide the sequence of steps required so that the method 'matching on the propensity score' leads toward identification of the parameter ATET. **[4 points]**

(g) Assume our interest are quantile effects (rather than mean effects) of the treated. Show the identification problem associated with the policy value '$F(Y^1 - Y^0 \mid D=1)$' and briefly and concisely compare this policy value to that described in (d). **[2 points]**

(h) Lets go back to thinking about ATET. I have succeeded in obtaining $AT\hat{E}T_N$ from matching using the random sample n. My next concerned is to make inference from the sample to the population relying fully on non-parametric bootstrap procedures. It is decided that I will use **asymptotic refinements** in order to elicit **asymmetric** critical values from the empirical distribution of the normalized variation in the estimated parameter $AT\hat{E}T_N$. How would you go about this? What would be the methodological difference if instead you were interested on a **symmetric** test? Be brief and concise in your answer.

[6 points]