University College London
Department of Economics
MECT2: Econometrics

May 2007: Time allowed: 2 hours

Answer all parts of Question 1 and any two of Questions 2 - 5.

Compulsory Question 1. Provide short answers to each of parts (a) - (j). [40 points; 4 points for each part]

(a). Given iid observations $\{(y_i, x_i) : i = 1, ..., n\}$, write down a kernel estimator for $E[y|x]$. Assume that y and x are both univariate.

(b). What role does the bandwidth play in kernel estimation? How does the choice of bandwidth effect the bias and variance of a kernel estimator?

(c). Give an example of an asymptotically pivotal statistic.

(d). Give two reasons why a researcher might choose to use the bootstrap for statistical inference.

(e). Consider a random sample of n observations $\{(y_i, z_i) : i = 1, ..., n\}$ where some observations of y_i are missing and $z_i = 0$ if y_i is missing, $z_i = 1$ if y_i is observed. The missing values of y_i are coded as zeros, though $z_i = 0$ in this case indicates that y_i is not observed. Specifically, for each observation i there is an unobserved random variable of interest $y_i^* \sim$Bernoulli(α) such that $y_i = z_i y_i^*$, i.e. the true value of y_i^* is observed only when $z_i = 1$. Derive sharp bounds on the variance of y_i^* as a function of the population distribution of (y_i, z_i).

(f). What are the restrictions on the censoring process implied by the Tobit model? Provide an economic interpretation of these restrictions.

(g). For a heterogeneous treatment effects model, define the average treatment on the treated parameter (ATT). When would it differ from the Average Treatment Effect?

(h). Explain the term 'selection on the observables'. Write down the two key assumptions in the Method of Matching for estimating the ATT and explain why they are important?

(i). Explain why the difference in differences estimator relies on a common trends assumption.

(j). Explain why an exclusion restriction is required for nonparametric identification in the selection model.

Question 2. [30 points]

Consider the following binary choice model:

$$\begin{aligned} Y^* &= X\beta + u, \\ Y &= 1\left[Y^* > 0\right], \end{aligned}$$

where X is a k-vector. Let $F(\cdot|x)$ denote the conditional distribution of u given $X = x$. Suppose a researcher has a random sample of observations of (Y, X), denoted $\{(y_i, x_i), i = 1, ..., n\}$ from which she wants to estimate β. Assume the standard rank condition that the support of X is not contained any proper linear subspace of $\mathbb{R}^k$.

(a) Suppose the researcher assumes that u is independent of X and is distributed $\mathcal{N}(0, 1)$. Then the Probit estimator can be used to estimate β. Write down the log-likelihood function for the Probit estimator.

(b) Now suppose that the researcher is concerned that the assumption that u is normally distributed may be incorrect. The researcher continues to assume that u is independent of X, but with an unknown distribution. She thus decides to estimate β using the maximum rank correlation estimator, $\hat{\beta}_{MRC}$, which is that value of $b \in \{\beta : \|\beta\| = 1\}$ that maximizes

$$\hat{S}(\beta) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1,\, j\neq 1}^{n} \left[1\{y_i > y_j\} 1\{x_i\beta > x_j\beta\} + 1\{y_i \leq y_j\} 1\{x_i\beta \leq x_j\beta\}\right].$$

(i) Why is the scale normalization imposed that $\hat{\beta}_{MRC}$ has Euclidean norm 1? Does the Probit model also make use of a scale normalization?

(ii) For identification of β (and consistency of $\hat{\beta}_{MRC}$), an additional assumption is required concerning the support of X. Provide a sufficient condition on the support of X that yields point identification of β.

(iii) Define $\beta^* \equiv \beta / \|\beta\|$. Given some additional regularity conditions, Sherman (1993, *Econometrica*) showed that $\sqrt{n}\left(\hat{\beta}_{MRC} - \beta^*\right)$ has an asymptotic normal distribution. Suppose that the regularity conditions are satisfied and that in addition u is normally distributed with variance 1. Of the Probit and MRC estimators, which is asymptotically more efficient?

(iv) Suppose that all the required assumptions for asymptotic normality of the MRC estimator hold, but that u is not normally distributed. Is the Probit estimator consistent?

(c) Discuss the differences between the parametric Probit estimator and the semi-parametric maximum rank correlation estimator. Give some advantages of each.

Question 3. [30 points]

Suppose that one wants to estimate the mean of a continuous distributed univariate random variable X with support $[0,1]$. However, X is observable with probability p, and is unobservable (missing) with probability $1-p$. The researcher's goal is to estimate a confidence interval for the population parameter $\theta_0 \equiv E[X]$. The researcher has a random sample of n observations of $\left(\tilde{X}, Z\right)$, denoted $\{(\tilde{x}_i, z_i) : i = 1, ..., n\}$ where $Z = 1$ if X is observed, and $Z = 0$ if X is unobservable. Thus, $\tilde{x}_i = z_i x_i$. Let $p = \Pr\{Z = 1\}$. p is not known to the researcher, but is assumed less than $1 - \epsilon$ for some $\epsilon > 0$ and greater than $\frac{1}{2}$. Assume that the variance of X and $\tilde{X}$ are strictly positive.

(a) Because of missing data θ_0 is only partially identified. Denote the identified set for θ_0 as $[\theta_l, \theta_u]$. Solve for θ_l and θ_u as functions of identified quantities.

(b) Formulate consistent estimators for θ_l and θ_u, denoted $\hat{\theta}_l$ and $\hat{\theta}_u$. Is it strictly greater than 0?

(c) What is the asymptotic distribution of $\sqrt{n}\begin{pmatrix} \hat{\theta}_l - \theta_l \\ \hat{\theta}_u - \theta_u \end{pmatrix}$? Justify your answer.

(d) Let $\hat{\sigma}_l$ and $\hat{\sigma}_u$ denote consistent estimators for the standard deviation of $\hat{\theta}_l$ and $\hat{\theta}_u$, respectively. What are

$$\lim_{n\to\infty} \Pr\left\{\theta_l \in \left[\hat{\theta}_l - \frac{\hat{\sigma}_l}{\sqrt{n}}C, \hat{\theta}_u + \frac{\hat{\sigma}_u}{\sqrt{n}}C\right]\right\},$$

and

$$\lim_{n\to\infty} \Pr\left\{\theta_u \in \left[\hat{\theta}_l - \frac{\hat{\sigma}_l}{\sqrt{n}}C, \hat{\theta}_u + \frac{\hat{\sigma}_u}{\sqrt{n}}C\right]\right\}$$

for fixed $C > 0$?

(e) Suppose that the population value θ_0 lies on the interior of the identified set, i.e. suppose $\theta_l < \theta_0 < \theta_u$. Show that for any $C > 0$.

$$\lim_{n\to\infty} \Pr\left\{\theta_0 \in \left[\hat{\theta}_l - \frac{\hat{\sigma}_l}{\sqrt{n}}C, \hat{\theta}_u + \frac{\hat{\sigma}_u}{\sqrt{n}}C\right]\right\} \geq \max_{\theta\in\{\theta_l,\theta_u\}} \lim_{n\to\infty} \Pr\left\{\theta \in \left[\hat{\theta}_l - \frac{\hat{\sigma}_l}{\sqrt{n}}C, \hat{\theta}_u + \frac{\hat{\sigma}_u}{\sqrt{n}}C\right]\right\}$$

(f) Suppose one wants to construct a confidence set for θ_0 with asymptotic coverage 0.95. What value of C would provide an asymptotic 0.95 confidence interval for $[\theta_l, \theta_u]$ such that

$$\lim_{n\to\infty} \left\{\theta_0 \in \left[\hat{\theta}_l - \frac{\hat{\sigma}_l}{\sqrt{n}}C, \hat{\theta}_u + \frac{\hat{\sigma}_u}{\sqrt{n}}C\right]\right\} \geq 0.95,$$

holding with equality for $\theta_0 = \theta_l$ or $\theta_0 = \theta u$?

Question 4. [30 points]

Consider the censored regression model where x_i are observable covariates and where observations on y_i^* are censored according to

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Assuming $u|x \sim \mathcal{N}(0, \sigma^2)$, show that the solution to the first-order conditions for ML estimation can be interpreted as an EM algorithm.

(b) Explain the difference between Random Sampling and Exogenous Stratification. Show that the Tobit ML estimator remains consistent under exogenous stratification.

(c) Outline and motivate the semiparametric censored least squares estimator for the coefficients β.

(d) Derive a test for the exogeneity of an explanatory variable. State clearly any further assumptions you make.

Question 5. [30 points]

Consider the binary treatment model:

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

where x_i are observable covariates and where the treatment indicator $d_i = 1(z_i'\gamma > v_i)$.

(a) Assuming u and v have a joint normal distribution, derive the conditional mean of u given $d = 1$. Show how this result can be used to formulate a Control Function estimator for the Average Treatment Effect. Relate this to the Heckman estimator for the selection model.

(b) Define an instrumental variable. Write down the conditions for an Instrumental Variables (IV) estimator to provide a consistent estimator of the ATE. In the context of a binary instrumental variable examine the difference between the LATE parameter and the ATE.