

University College London
Department of Economics
MECT2: Econometrics

22nd May 2006: 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). When is a test said to be consistent? Define the consistency of a test.
- (b). Explain briefly the likelihood ratio test in the setup of maximum likelihood estimation.
- (c). Give an example of asymptotically pivotal test statistics.
- (d). Explain briefly the curse of dimensionality in nonparametric estimation.
- (e). Choosing a good bandwidth is important for kernel density estimation. What is a bias-variance tradeoff when one tries to choose a bandwidth? Explain this briefly.
- (f). Explain briefly how the selection model relaxes the assumptions in the censored regression model.
- (g). What is the difference between the homogeneous and heterogeneous treatment effects model?
- (h). Explain the terms ‘selection on the observables’ and ‘selection on the unobservables’.
- (i). Define the average treatment on the treated parameter (ATT). Write down the two key assumptions in the Method of Matching for estimating the ATT and explain why they are important.
- (j). What characterizes a good discontinuity design?

Question 2. [30 points]

Let $K(\cdot)$ denote a kernel function and let $h > 0$ denote a bandwidth. Suppose that we wish to estimate the derivative $f'(x)$ of an unknown function $f(x)$. Suppose that a nonparametric estimator $\hat{f}(x)$ of $f(x)$ is available, but it is not smooth. As a result, $\hat{f}'(x)$ is not well-defined. To overcome this problem, one can use a smoothed estimator of $\hat{f}'(x)$, that is

$$\tilde{f}(x) = \int_{-\infty}^{\infty} \hat{f}(s) \frac{1}{h} K\left(\frac{x-s}{h}\right) ds$$

with a differentiable K . Assume that $\hat{f}(x)$ is an unbiased estimator of $f(x)$, hence $E[\hat{f}(x)] = f(x)$.

(a) Make some assumptions on f and K and show that $\lim_{n \rightarrow \infty} E[\tilde{f}(x)] = f(x)$ for each x .

(b) Make some assumptions on f and K and show that $\lim_{n \rightarrow \infty} E[\tilde{f}'(x)] = f'(x)$ for each x .

(c) Suppose that $f(x)$ is the cumulative distribution function (CDF) of a continuous random variable and $\hat{f}(x)$ is the corresponding empirical CDF. Obtain the explicit form of $\tilde{f}'(x)$ in this case.

Question 3. [30 points]

Richard Schmalensee and Thomas Stoker (1999, *Econometrica*) consider estimation of household gasoline demand in the US using household survey data. The dependent variable (Y) is log gallons of gasoline consumed and independent variables are log household income (X_1), log ages of the head of the household (X_2), and a vector of discrete variables (Z) for household characteristics including the number of drivers, household size, location, and so on. The variables X_1 and X_2 are continuously distributed.

(a) Consider fully nonparametric estimation of household demand and specify the econometric model accordingly. Discuss potential problems with this approach.

(b) Consider partially linear estimation of household demand and suggest a specification of the econometric model. Explain why your specification can be called a partially linear model.

(c) Discuss the advantages and disadvantages of your specification in part (b).

(d) Suppose that you wish to learn about the income elasticity. In your specification in part (b), how would you define the income elasticity? It is believed that the income elasticity can depend on the level of the income. Is your specification plausible given this belief?

(e) Suppose that people tend to choose a longer commute to a job that pays a higher salary. Is there any problem with this in terms of interpretation of your estimation results?

Question 4. [30 points]

Consider the normal linear latent model

$$y_i^* = x_i' \beta + u_i \text{ where } u_i \sim N(0, \sigma^2)$$

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. Suppose you wish to assess whether the distribution of u_i is homoskedastic. You are given a set of observable covariates z of dimension l that are thought to determine the heteroskedasticity in the distribution of u_i . Write down an expression for the score test of homoskedasticity.

b. Derive a test for the exogeneity of an explanatory variable where you can assume the unobservables in the underlying structural model and jointly normally distributed. State clearly any further assumptions you make.

c. Suppose you want to estimate β and suppose that you are willing to assume that u_i has a symmetric although not necessarily normal distribution. Write down and interpret an appropriate semiparametric estimator.

Question 5. [30 points]

1. Consider the treatment response model:

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

where y_i is the outcome variable, X_i are observable covariates, u_i is an unobservable determinant of y_i , α_i measure the response of individual i to a binary treatment d_i with assignment rule:

$$d_i = 1(Z_i'\gamma + v_i > 0)$$

with Z_i describing the observable determinants of assignment to treatment and v_i the unobservable determinants.

In the context of this model:

a. Describe the assumptions under which OLS will recover the average treatment effect (ATE). Show that the OLS is a consistent estimator of the ATE under those conditions.

b. Define an instrumental variable. Write down the conditions for an Instrumental Variables (IV) estimator to provide a consistent estimator of the ATE.

c. For the case of a binary instrumental variable examine the difference between the Local Average Treatment Effect (LATE) parameter and the ATE. You may assume that there are no other covariates in the regression model.

d. Contrast the identifying assumption for Discontinuity Design with those for LATE.