UCLA Economics

Spring 2002 Quantitative Methods Comprehensive Examination

There are three sections. Please answer all of the questions; use a separate blue book for each of the three sections. For your information, the 0.90 and 0.95 quantiles of the Chi-squared distribution with one degree of freedom are 2.78 and 3.84.

Part 1

1) Let X be a binary variable with $\Pr(X = 1) = \Pr(X = 0) = ½$. Conditional on $X = x$, the random variable Y has an exponential distribution with arrival rate $1 + \theta x$. (Note: the probability density function for an exponential random variable Z with arrival rate λ is $f_Z(z;\lambda) = \lambda \exp(-z\lambda)$.)

   a) What is the mean of Y?

   b) What is the probability that $X = 1$ given $Y \leq 1$?

   c) Suppose that $(X_1, Y_1), \ldots, (X_N, Y_N)$ are a random sample from this distribution. What is the maximum likelihood estimator for θ?

   d) What is the large sample variance of the maximum likelihood estimator?

   e) Construct a moment estimator for θ based on the expected value of Y. What do you expect the variance of this estimator to be relative to the variance of the maximum likelihood estimator? There is no need to actually calculate this variance.

2) Let $X_1, \ldots, X_N$ be a random sample from a normal distribution with mean zero and variance σ^2.

   a) What is the Cramér-Rao bound for unbiased estimators for σ^2, given $N = 1000$?

   b) Let $N = 1000$, and $\sum_{i=1}^{1000} X_i = 10$ and $\sum_{i=1}^{1000} X_i^2 = 1040$. Test the null hypothesis that $\sigma^2 = 1$ against the alternative that $\sigma^2 \neq 1$ at the 5% level using a likelihood ratio test.

   c) Carry out the same test using a Lagrange multiplier test.

   d) Carry out the same test using a Wald test.

   You can use large sample approximation to the distribution of all the test statistics.

Part 2

3) Consider the linear model

$$y_i = \beta_1 + \beta_2 x_i + u_i, \qquad i = 1, \ldots, n \tag{2.1}$$

   where $\{(y_i, x_i)\}$ is an independent and identically distributed sequence and $E\{y_1 | x_1\} = \beta_1 + \beta_2 x_1$ a.s..

   a) Suppose one tries to test $\beta_2 = 0$ versus $\beta_2 \neq 0$ using a t-test.

i) Using either a graph or a mathematical derivation, illustrate how the probability of making a type II error changes with the sample size.

ii) How does the probability of making a type I error change with the sample size?

b) Suppose you are not certain that x_i is exogenous. Explain **briefly** how you would test the exogeneity of x_i here.

4) Consider the linear regression models

$$y_i = x_i\beta + v_i, \tag{2.2}$$

$$y_i = x_i\beta + z_i'\gamma + u_i, \qquad i = 1, \dots, n \tag{2.3}$$

where $\{[y_i, x_i, z_i']'\}$ is an independent and identically distributed sequence of random vectors for which $E\{u_1 | x_1, z_1\} = 0$ a.s.. (Note that x_1 is a scalar.) Denote the ordinary least squares estimator of β in (2.3) by $\hat{\beta}_L$ and that of β in (2.2) by $\hat{\beta}_S$.

a) Suppose that $\gamma \neq 0$ and z_1 is correlated (but not collinear) with x_1 and that the purpose is to estimate β accurately. Is $\hat{\beta}_L$ necessarily a better estimator than $\hat{\beta}_S$? Explain. (No derivations needed.)

b) Show that $\hat{\beta}_L = (X'MX)^{-1} X'My$ with $M = I - Z(Z'Z)^{-1} Z'$, where I is the identity matrix.

c) Show that $\hat{\beta}_L$ can alternatively be computed by regressing y_i on the residuals of a regression of x_i on z_i.

5) Consider the linear regression model

$$y_i = x_i\beta + z_i\gamma + u_i, \qquad i = 1, \dots, n \tag{2.4}$$

where $\{(y_i, x_i, z_i)\}$ is an independent and identically distributed sequence, x_i is independent of z_i and both x_i and z_i are dummy variables. Assume that $E\{u_1 | x_1, z_1\} = 0$ a.s.. Suppose that $\text{var}(u_1 | x_1 = 0, z_1 = 0) = 2$ and suppose that $\text{var}(u_1 | x_1, z_1) = 1$ for all other possible combinations of x_i, z_i. Suppose further that $\Pr(x_1 = 1) = \Pr(z_1 = 1) = 1/2$.

a) Show that the (asymptotic) OLS variance matrix is

$$\frac{4}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}. \tag{2.5}$$

b) It can be shown that $\text{var}(u_1 | x_1) = (3 - x_1)/2$ a.s.. Suppose that we estimate the model

$$y_i^* = x_i^*\beta + z_i^*\gamma + u_i^*, \qquad i = 1, \dots, n \tag{2.6}$$

with $y_i^* = w_i y_i$, $x_i^* = w_i x_i$, $z_i^* = w_i z_i$, and $u_i^* = w_i u_i$ where

$$w_i = \frac{1}{\sqrt{(3-x_i)/2}}. \tag{2.7}$$

The asymptotic variance matrix of the OLS estimator in (2.6) is

$$\begin{bmatrix} 6 & -8 \\ -8 & 16 \end{bmatrix}. \tag{2.8}$$

Explain how it can be true that the variances of the OLS estimator in the transformed model exceed those in the original model even though we have used the known conditional variances $\operatorname{var}(u_i | x_i)$ to correct for heteroskedasticity.

Part 3

6) Let $(y_{1i}^*, y_{2i}^*, y_{3i}^*)$ be a three-dimensional vector of continuous random variables that are independent across $i = 1, 2, \ldots, n$. These random variables are not observed. Instead, we observe z_i and y_i defined as follows:

$$z_i = \begin{cases} y_{2i}^* & \text{if } y_{1i}^* > 0 \\ y_{3i}^* & \text{if } y_{1i}^* \le 0 \end{cases}$$

and

$$\begin{aligned} y_i &= \begin{cases} y_{1i}^* & \text{with probability } \gamma \\ 0 & \text{with probability } 1-\gamma \end{cases} \quad \text{if } y_{1i}^* > 0 \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{if } y_{1i}^* \le 0 \end{aligned}.$$

The parameters to be estimated include γ and any parameters of the joint density for $(y_{1i}^*, y_{2i}^*, y_{3i}^*)$. Write down the likelihood function.

7) Suppose you observe wage w_i and individual characteristics x_i for n randomly selected individuals. Throughout the problem, assume that n is large (greater than 1,000).

   a) Assume that x_i is one-dimensional, binary random variable, i.e. $x_i \in \{0, 1\}$. Write down a kernel regression estimator of the wage given x_i.

   b) Assume that x_i is two dimensional with one continuous component and the other a binary random variable. Write down a kernel regression estimator of the wage given x_i. How do you select the bandwidth(s)?

   c) What is the effect of random measurement error in x_i on the regression estimator in (b)?