ECONOMETRICS - QUALIFIER

May 2007

ANSWER ONLY 5 QUESTIONS.

1. Let X be the female dummy variable ($X = 1$ for female and $X = 0$ for male) Y denote starting salary (in \$1,000) of an individual working in firm A. The joint distribution of gender and the starting salary of an individual in firm A is given by

Table1: Joint Probability Function $f_{XY}(x, y)$ for Firm A

X/Y	40	50	60
0	0.3	0.2	0.2
1	0.1	0.1	0.1

   (a) Compute the conditional probability functions for female and male, i.e., compute $f\{_{Y|X}\}(y|x = 1)$ and $f_{Y|X}(y|x = 0)$.

   (b) Are X and Y independent? Explain why of why not.

   (c) Bob claims that firm A discriminates against female because more male making \$60,000 than female (i.e., $f_{XY}(0, 60) = 0.2 > f_{XY}(1, 60) = 0.1$). Assuming that all the new hires of firm A have the same level of education,discuss whether you agree with Bob's argument or not (give your reasonings).

2. Suppose that y_i given x_i has a Poisson distribution with probability

$$f(y_i|x_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \qquad y_i = 0, 1, 2, 3, ...$$

where $\lambda_i = \lambda_{x_i}$. Assume that $\lambda_{xi} = exp(x_i'\beta)$, where x_i is $k\times$, β is a $k \times 1$ vector of parameter.

   (a) Discuss how to estimate β by the maximum likelihood estimation method. You need to set up the log-likelihood function, and derive the first order condition.

   (b) Suppose that y_i = number of clinic visits in a given year for individual i, $x_i = (1, gender_i,$ age_i, $income_i)$, with estimated coefficients given by $(\widehat{\beta} = \widehat{\beta}_1, ..., \widehat{\beta}_4)\prime$. Discuss how can you test for the hypothesis that the number of clinic visits is not related to gender, against the alternative hypothesis that male tend to visit clinic more often than female. Using a 5% level test, state the distribution of your test statistic and the critical value.

   (c) Discuss how can you test the joint null hypothesis that gender and income do not affect the number of clinic visits.

3. Let

$$X_i = X + \xi_i \quad \text{and} \quad Y_i = 1 + \eta_i,$$

where $(\xi_i, \eta_i)'$ is an iid random vector with mean zero and covariance matrix given by

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix},$$

and X is a random variable such that $\text{var}(X) \neq 0$ and it is independent of $(\xi_i, \eta_i)'$ for all $i \geq 1$. Moreover, we define $\bar{X}_n$ and $\bar{Y}_n$ to be

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i,$$

i.e., the sample means of (X_i) and (Y_i) respectively.

Answer the following questions:

(a) State and verify whether the law of large numbers hold for each of (X_i) and (Y_i).

(b) Let $\bar{Z}_n = (\bar{X}_n, \bar{Y}_n)'$. Define a sequence (c_n) of sample size n and a random vector Z so that we have

$$c_n(\bar{Z}_n - Z) \to_d \mathbb{N}(0, \Sigma)$$

for some $\Sigma > 0$ as $n \to \infty$. Specify (c_n), Z and Σ.

(c) Show that

$$\frac{\bar{X}_n}{\bar{Y}_n} = X + O_p(n^{-1/2}),$$

for large n.

(d) Show that, as $n \to \infty$,

$$\sqrt{n}\left(\frac{\bar{X}_n}{\bar{Y}_n} - X\right) \to_d \sqrt{1 - 2X + 2X^2}\,\mathbb{N}(0, 1),$$

where $\mathbb{N}(0, 1)$ is a standard normal random variate independent of X.

4. Consider the regression model

$$y_i = w_i'\alpha + x_i'\beta + u_i,$$

where (w_i) and (x_i) are the regressors and (u_i) are the regression errors for $i = 1, \ldots, n$.

Answer the following questions:

(a) Let (y_i^*) and (x_i^*) be the fitted residuals respectively from the regressions of (y_i) and (x_i) on (w_i), and consider the regression

$$y_i^* \quad \text{against} \quad x_i^*.$$

Are the fitted residuals from this regression the same as those from the original regression? What about R^2's?

(b) Consider the regression

$$y_i \quad \text{against} \quad w_i \ \text{ and } \ w_i + x_i.$$

Show that the F statistic on the coefficient of $(w_i + x_i)$ in this regression is numerically identical to that on the coefficient of (x_i) in the original regression.

(c) Consider the step-wise OLS regressions, i.e.,

$$y_i = w_i'\tilde{\alpha} + \tilde{v}_i$$

and

$$\tilde{v}_i = x_i'\tilde{\beta} + \tilde{u}_i.$$

Assume that the standard assumptions required for the asymptotics of the classical regression model hold. Find the probability limit of $\tilde{\beta}$ and compare it with the OLS estimator $\hat{\beta}$ of β in the original regression. Is $\tilde{\beta}$ consistent? If not in general, provide extra conditions under which $\tilde{\beta}$ is consistent.

(d) Now we assume that (w_i) and (x_i) are independent each other and that

$$\frac{1}{n}\sum_{i=1}^{n} w_i x_i' \to_p 0 \quad \text{and} \quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i x_i' = O_p(1)$$

as $n \to \infty$ in question (c) above. Show that $\tilde{\beta}$ is consistent in this case. Find the limit distribution of $\tilde{\beta}$. Is it identical to that of $\hat{\beta}$ from the original regression?

5. Consider an identified simultaneous equation

$$y_1 = Y_2\gamma + X_1\beta_1 + u_1$$

with its reduced form equations

$$y_1 = X_1\pi_{11} + X_2\pi_{21} + v_1, \qquad Y_2 = X_1\Pi_{12} + X_2\Pi_{22} + V_2$$

where rows (v_{t1}, V_{t2}) are temporally independent, identically distributed multivariate normal with zero mean and covariance $\Omega = (\Omega_{ij})$, y_1 and Y_2 are endogenous variables, and X_1 and X_2 are exogenous variables. The relationships between the structural and reduced form parameters are

$$\pi_{11} - \Pi_{12}\gamma = \beta_1, \qquad \pi_{21} - \Pi_{22}\gamma = 0, \qquad v_1 - V_2\gamma = u_1$$

(a) It is often said that the first equation is identified if structural coefficients γ and $\beta_1$can be recovered from the estimates of reduced form coefficients $\widehat{\pi}_{i1}$and $\widehat{\Pi}_{i2}$. What are the conditions on the reduced form coefficients to make the first equation to be identifiable?

(b) We wish to test the null hypothesis H_0: $cov(u_1, Y_2) = 0$. Show how to compute a test statistic and present the theory behind it.

(c) Derive the IV estimator of the structural coefficients and its asymptotic covariance matrix.

(d) Derive the efficient GMM estimator of the structural coefficients.

6. Consider two regression equation model of latent variables

$$\begin{aligned} y_{t1}^* &= X_{t1}\beta_1 + u_{t1} \\ y_{t2}^* &= X_{t2}\beta_2 + u_{t2} \end{aligned}$$

where the first equation is the relationship of primary interest, and the second equation is the sample selection equation. The sample selection is determined by

$$y_{t1} = y_{t1}^* \qquad if \qquad y_{t2}^* > 0$$

and no observations on y_{t1}^* and X_{t1} if $y_2^* \leq 0$. Error vector $(u_{t1}, u_{t2})'$ are i.i.d. normal with zero mean vector and covariance matrix $\sum = (\sigma_{ij})$ with a nonzero covariance.

(a) Derive the mean of observed y_{t1}, i.e., $E(y_{t1})$, and explain why the OLS estimator $\widehat{\beta}_1$ from a regression of observed y_{t1} on observed X_{t1} is biased.

(b) Explain Heckman's idea of estimating β_1 based on the specification of $E(y_{t1})$.

(c) Specify the likelihood function.
Hint: Let $X \sim N(\mu, \sigma^2)$. Then,

$$E(X|X > c) = \mu + \sigma\frac{\phi(b)}{1 - \Phi(b)}, \qquad E(X|X \leq c) = \mu - \sigma\frac{\phi(b)}{\Phi(b)}, \qquad b = (c - \mu)/\sigma$$

Hint: Let z_1 and z_2 be jointly normal with means μ_1 and μ_2, and covariance matrix $\sum = (\sigma_{ij})$. Then, the conditional distribution of z_1 given z_2 is normal

$$z_1|z_2 \sim N[\mu_1 + (\sigma_{12}/\sigma_{22})(z_2 - \mu_2), \qquad \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{22}]$$