PROBLEM SET 7

Problem 1 (Probit Selection Model)

Consider the following Probit selection model:

$$\begin{aligned} Y_1 &= Y_2 \cdot Y_1^* \\ Y_1^* &= X_1\beta_1 + U_1, \\ Y_2 &= 1\left[Y_2^* \geq 0\right], \\ Y_2^* &= X_2\beta_2 + U_2, \end{aligned}$$

from which one obtains a random sample of observations of (Y_1, Y_2, X_1, X_2), denoted $\{(y_{1i}, y_{2i}, x_{1i}, x_{2i}) : i$ where Y_1^* is the outcome variable of interest and the realization of Y_2 determines whether $Y_1 = Y_1^*$ or $Y_1 = 0$. Assume that (U_1, U_2) and $X = (X_1, X_2)$ are independent, and that (U_1, U_2) are bivariate normally distributed, each with mean zero, variances $var(U_1) = var(U_2) = 1$, and covariance $cov(U_1, U_2) = \rho$

Q1 What is the selection probability $\Pr\{Y_2 = 1|X = x\}$? How can you consistently estimate the parameters of this equation?

Answer $\Pr\{Y_2 = 1|X = x\} = \Phi(x_2\beta_2)$. β_2 is consistently estimated by running a probit regression of Y_2 on X_2, as long as $E[X_2'X_2]$ is invertible.

Q2 Solve for the conditional density for Y_1 given $Y_2 = 1$ and X. What is $E[Y_1|Y_2 = 1, X = x]$?

Hint: By joint normality of U_1 and U_2, the conditional distribution of U_1 given $U_2 > c$ can be deduced as a function of c and ρ as

$$g(u_1; c, \rho) = f(u_1|U_2 > c) = \frac{1}{1 - \Phi(c)} \int_c^\infty f(u_1|U_2 = u_2)\, \phi(u_2)\, du_2,$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal pdf and cdf, respectively. You may use the function $g(u_1; c, \rho)$ in your answer.

Furthermore, recall that the distribution of U_1 conditional on $U_2 = c$ is

$$U_1|\left(U_2 = c\right) \sim \mathcal{N}\left(\rho c, 1 - \rho^2\right),$$

with conditional pdf

$$f_{U_1|U_2}(u_1|U_2 = c) = \frac{1}{\sqrt{1-\rho^2}}\phi\left(\frac{u_1 - \rho c}{\sqrt{1-\rho^2}}\right),$$

i.e. the pdf of a normal random variable with mean ρc and variance $1 - \rho^2$. Because U_1 is normally distributed with mean zero and unit variance,

$$E[U_1|U_1 > c] = \frac{\phi(c)}{1 - \Phi(c)}.$$

Answer Conditional on $Y_2 = 1$, we have $Y_1 = X_1\beta_1 + U_1$. The conditional distribution of Y_1 is therefore

$$Y_1|(Y_2 = 1, x) \sim x_1\beta_1 + U_1|(Y_2 = 1, X = x).$$

The distribution of U_1 given $(X = x, Y_2 = 1)$ is that of $U_1|U_2 > -x_2\beta_2$. Therefore the density of Y_1 given $(X = x, Y_2 = 1)$ is

$$f_{Y_1|X,Y_2}(y_1|x, 1) = g(y_1 - x_1\beta_1; -x_2\beta_2, \rho),$$

which is that of $U_1|U_2 > -x_2\beta_2$ shifted by $x_1\beta_1$. Furthermore

$$\begin{aligned} E[Y_1|Y_2 = 1, X = x] &= E[Y_1|U_2 > -x_2\beta_2, X = x] \\ &= x_1\beta_1 + E[U_1|U_2 > -x_2\beta_2], \end{aligned}$$

where

$$\begin{aligned} E[U_1|U_2 > -x_2\beta_2] &= \frac{1}{\Pr\{U_2 > -x_2\beta_2\}}\int_{-x_2\beta_2}^{\infty} E[U_1|U_2 = u_2]\,\phi(u_2)\,du_2 \\ &= \frac{1}{\Phi(x_2\beta_2)}\int_{-x_2\beta_2}^{\infty} \rho u_2\phi(u_2)\,du_2 \\ &= \rho\cdot\left(\frac{1}{\Phi(x_2\beta_2)}\int_{-x_2\beta_2}^{\infty} u_2\phi(u_2)\,du_2\right) \\ &= \rho\cdot E[U_2|U_2 > -x_2\beta_2] \end{aligned}$$

Thus $E[Y_1|Y_2 = 1, X = x] = x_1\beta_1 + \rho\lambda(-x_2\beta_2)$ where $\lambda(c) = \dfrac{\phi(c)}{1 - \Phi(c)}$.

Q3 What is the log likelihood for a random sample of observations of (Y_1, Y_2, X_1, X_2)?

Answer

$$\begin{aligned} l(\beta, \rho) &= \sum_{i=1}^{n} \log\left(\Pr\{Y_2 = 0|X = x_i\}^{1-y_{i2}}\cdot\left[f_{Y_1|X,Y_2}(y_1|x_i, 1)\Pr\{Y_2 = 1|X = x_i\}\right]^{y_{i2}}\right) \\ &= \sum_{i=1}^{n}(1 - y_{i2})\log[1 - \Phi(x_{i2}\beta_2)] + y_{i2}\log\{g(y_{i1} - x_{i1}\beta_1; -x_{i2}\beta_2, \rho)\cdot\Phi(x_{i2}\beta_2)\}. \end{aligned}$$

Q4 Propose two different ways to consistently estimate β_1. Which provides a more efficient estimator asymptotically?

Answer (i) A two stage procedure. First run a probit regression of y_{i2} on x_{i2} to consistently estimate β_2. Then estimate $E[Y_1|Y_2 = 1, X = x] = x_1\beta_1 + \rho\lambda(-x_2\beta_2)$ in a second stage linear mean regression of y_{i1} on x_{i1} and $\lambda\left(-x_{i2}\hat{\beta}_2\right)$ to produce consistent estimates for β_1 and ρ.

(ii) Maximum likelihood: maximize the likelihood from Q3 above. Maximum likelihood will provide an asymptotically efficient estimator.

Q5 Are the parameters β_1, β_2, ρ identified? What further conditions do you need to impose, if any?

Answer Possible answer 1: From the two stage procedure of Q4 we see that invertibility of $E[X_2'X_2]$ is sufficient for identification of β_1. Invertibility of $E\left[(X_1, \lambda(-x_2\beta_2))'(X_1, \lambda(-x_2\beta_2))\right]$ is sufficient for identification of β_2 and ρ.

Possible answer 2: Sufficient conditions for $E\left[\frac{1}{n}l(\beta, \rho)\right]$ to be concave in (β, ρ) will guarantee identification.

Q6 Why might one wish to test the hypothesis that $\rho = 0$? What implication would this have?

Answer $\rho = 0$ would mean that U_1 and U_2 are independent conditional on X. Therefore selection/censoring is exogenous conditional on X and one may consistently estimate β_1 simply by running a linear regression of the non-censored observations of Y_1 on X_1 since in this case $E[Y_1|Y_2 = 1, X = x] = x_1\beta_1$.