PROBLEM SET 5

Problem 1 (The type I Tobit model with neglected heterogeneity)

Consider the following model of censoring

$$Y^* = X\beta + W\gamma + \epsilon,$$
$$Y = \max(Y^*, 0)$$

where $(X, W) \perp \epsilon$, $X \perp W$, and $\epsilon \sim \mathcal{N}(0, 1)$, where $\perp$ denotes stochastic independence.

(a) Suppose a random sample of observations of (Y, X, W) is observed. Can the parameters β, γ be consistently estimated, and if so, how?

Solution: This is a Tobit Model with covariates (X, W). The parameters can be consistently estimated by maximum likelihood.

(b) Suppose now that W is not observed, but rather only a random sample of (Y, X) is observed. Assume that $W \sim \mathcal{N}\left(0, \sigma_w^2\right)$. This is now a model of censoring with unobserved heterogeneity (W is the unobserved heterogeneity). Can β be consistently estimated in this model? Explain.

Solution: Define $\tilde{\epsilon} \equiv W\gamma + \epsilon$, so that $Y^* = X\beta + \tilde{\epsilon}$. The independence assumptions and normality of ϵ, W imply that

$$\tilde{\epsilon}|x \sim \mathcal{N}\left(0, \tilde{\sigma}^2\right),$$

where $\tilde{\sigma}^2 = \gamma^2\sigma_w^2 + 1$. Thus

$$Y^* = X\beta + \tilde{\epsilon},$$
$$Y = \max(Y^*, 0),$$

which is again the Tobit model, and β can be consistently estimated by maximum likelihood.

Problem 2 (A variation of the Tobit model)

Consider the model:

$$Y = \begin{cases} Y^* & \text{if } 0 < Y^* \leq 1 \\ 0 & \text{if } Y^* \leq 0 \\ 1 & \text{if } Y^* > 1 \end{cases},$$

$$Y^* = X\beta + U,$$

where U and X are independent and $U \sim \mathcal{N}\left(0, \sigma^2\right)$.

(a) Write the log-likelihood of an iid sample $\{(y_i, x_i) :, i = 1, ..., N\}$.

Solution:

$$L(\beta, \sigma) = \frac{1}{N} \sum_{i=1}^{N} \log \Phi \left(-\frac{x_i\beta}{\sigma}\right)^{1[y_i=0]} \log \Phi \left(\frac{x_i\beta - 1}{\sigma}\right)^{1[y_i=1]} \left[\log \sigma - \left(\frac{y_i - x_i\beta}{\sqrt{2}\sigma}\right)^2\right]^{1[0<y_i<1]}$$

(b) What are the asymptotic properties of the Maximum Likelihood estimator?

Solution: All the usual ML results apply, the parameter estimates are consistent, asymptotically normal, and attain the efficiency bound.

(c) How could you perform a specification test of this model?

Solution: Absolutely. White's ML specification test comparing the Hessian and the outer product for of the information matrix applies.

Problem 4 (RECAP + Probit Selection Model) [Is not being presented in tutorial]

i. [4] γ can be consistently estimated by the OLS estimator

$$\hat{\gamma}_{OLS} = \left(\sum_{i=1}^{n} x_i' x_i \right)^{-1} \sum_{i=1}^{n} x_i' y_{2i}$$

as long as $E[X'X]$ is nonsingular and $E[X'U_2] = 0$.

ii. [4] The IV/SLS estimator can be used as long as the necessary rank condition holds. For example, if X_2 is univariate (or if all but one component are dropped) and $E[X'(X_1, Y_2)]$ has full column rank then the IV/SLS estimator

$$\hat{\theta}_{IV} = \left(\sum_{i=1}^{n} x_i' (x_{i1}, y_2) \right)^{-1} \sum_{i=1}^{n} x_i' y_{1i}$$

is consistent for $\theta \equiv (\beta', \delta')'$.

i. [5] The probit (maximum likelihood estimator) is consistent, with the rank condition that $E[X'X]$ is nonsingular and the normalization that $\Omega_{22} = 1$ are required for identification.

$$\hat{\gamma}_{probit} = \arg\max_{\gamma} = \frac{1}{n} \sum_{i=1}^{n} (1 - y_i) \log (1 - \Phi(x_i \gamma)) + y_i \log (\Phi(x_i \gamma)).$$

ii. [4] Same as part a(ii) of this question: If X_2 is univariate and if $E[X'(X_1, Y_2)]$ has full column rank then the IV/SLS estimator

$$\hat{\theta}_{IV} = \left(\sum_{i=1}^{n} x_i' (x_{i1}, y_2) \right)^{-1} \sum_{i=1}^{n} x_i' y_{1i}$$

is consistent for $\theta \equiv (\beta', \delta')'$. Estimation could also be done by maximum likelihood or more generally via GMM.

i. [3] Same as part a(i) of this question, the OLS estimator is consistent under the usual rank condition and will here coincide with the ML estimator.

ii. [6] The parameters can be estimated by maximum likelihood, GMM, or a two-stage procedure that first estimates γ and then uses the first stage estimator to construct a control function in the estimation of the first equation as described in Wooldridge Section 15.7. The last stage will require a rank condition that $E[W'W]$ be non-singular, where $W = [X_1, Y_2, U_2]$.