TOBIT

$$y_i^* = x_i^T\beta + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$$

$$y_i = \begin{cases} 0 & y_i^* \le 0 \\ y & y_i^* > 0 \end{cases}$$

$$E(y_i^* \mid x_i) = x_i^T\beta$$

$$E(y_i \mid x_i) = \Phi\left(\frac{x_i'\beta}{\sigma}\right)(x_i'\beta + \sigma\lambda_i)$$

use this to describe random draws from pop'l that may or may not be censored.

note, it pertains to the observed y_i not the latent variable y_i^*

ex. we don't know whether recent will sell out or NOT

Marginal Effects

$$\frac{\partial E(y_i^* \mid x_i)}{\partial x_i} = \beta$$

$$\frac{\partial E(y_i \mid x_i)}{\partial x_i} = \beta\,\Phi\left(\frac{x_i'\beta}{\sigma}\right) \quad \text{It is attenuated}$$

NOTE: case (Censoring at Zero)

Specification issues

(1) Heteroskedasticity:

Tobit inconsistent and the degree depends on the extent of ~~s~~ censoring.

(2) nonnormality -

inconsistent.

Cragg Model (1971)

This is closely related to Tobit. However, there are two decisions being made and each is independent of the other.

(1) Should I buy a car?

~~(2)~~ NO → 0

yes → How much do I spend.

The threshold (index) is not leading to yes or no, only how much.

Probit

$$D_t^* = x_{1t}^T\beta_1 + \varepsilon_{1t}$$

$$D_t = \begin{cases} 1 & \text{if } D_t^* \geq 0 \\ 0 & \text{if } D_t^* < 0 \end{cases}$$

D_t^* might be "net benefits" assoc. with something unobserved. Through Revealed Preference. we assume a rational decisionmaker will choose to do something if net benefits ≥ 0.

Let A_t^* be the desired amount of some good. We only observe the actual amount A_t

$$\underline{A_t^* = x_{2t}^T\beta_2 + \varepsilon_{2t}}$$

$$A_t = \begin{cases} A_t^* & \text{if } A_t^* > 0 \\ 0 & \text{if not} \end{cases}$$

This is Basically Tobit.
note that the decision to use and the amount used are implicitly defined by the same mechanism. If $X_{2t}'\beta + \varepsilon_{2t}$ doesn't exceed some threshold (0) then you decide not to do ^any of the Action.

Double Hurdle (Cragg)

$$A_t = \begin{cases} A_t^* & \text{if } A_t^* > 0 \text{ AND } D_t^* > 0 \\ 0 & \text{if } D_t^* < 0 \end{cases}$$

The decisionmaker only chooses to use the item (A_t) if he desires ($D_t > 0$) and, if the desired amount, conditional on $X_{2t}\beta_2 + \varepsilon_{2t}$ is positive.

useful if $X_{1t} \neq X_{2t}$ or if the magnitudes $\beta_1 \neq \beta_2$ for some X_t.

ε_{1t} & ε_{2t} are NOT correlated

Cragg (Two steps)

1) Estimate Probit and compute inverse mills Ratio. $\frac{\phi(X_{1i}'\beta_1/\sigma)}{\Phi(X_{1i}'\beta_1/\sigma)}$

2) add IMR to a Truncated Regression

$\Rightarrow$ Decision to do something is independent of the decision about how much to do.

Example: Foreign Exchange derivatives yes or No.

~~If you use them, then~~

Those who use them must decide How much to use.

Heckit (Incidental Truncation)

In this case the errors of the Decision & Action are correlated. The decisions are actually related by some unobservable factors.

~~hours worked~~
~~hrs/wages) LFP = f(exper, age, educ, kids, ...)~~
~~(Decision to work or not (LFP))~~

LABOR Supply

Wage Eq (market wage - res wage)

we only observe whether you are in or out.

Hours Worked = f(kids, educ, other stuff)

you only observe Hours > 0 if

Market Wage > Res wage

If unobserved factors that affect these are correlated - Incidentally ~~Treated~~ truncated.

(If not - Cragg)

Two Estimators

1. Heckit – Two step.

(a) Estimate selection Eq using Probit

(b) form IMR and include it in an OLS Regression for the extent of use eq.

(c) Estimate the correct covariance matrix. (several possibilities) See Hill & Adams for details.

2. MLE. – Probably Better, esp if errors are normally dist.

Murphy Topel

Asy cov of 2 step MLE.

model 1 $E(y_1 \mid x_1, \theta_1)$ — from $\ln(L_1)$

model 2 $E(y_2 \mid x_2, \theta_2, E(y_1 \mid x_1, \theta_1))$ — from $\ln(L_2)$

Say you've got a likelihood function of 2 R.V. and 2 sets of parameters

$$\ln L = \sum f(y_{i1}, y_{i2} \mid x_{i1}, x_{i2}, \theta_1, \theta_2)$$

One way to proceed. Specify the entire joint density, f, and use FIML

OR do 2-steps.

Estimate model 1 $\Rightarrow \hat{\theta}_1, \hat{V}_1$ MLE and est of its Asy cov.

Sub predicted values into model 2

$$\ln \hat{L} = \sum f(y_{i2} \mid x_{i2}, \theta_2, (x_{i1}, \hat{\theta}_1))$$

maximize this w/respect to θ_2

$\hat{\theta}_2$ and $\hat{V}_2$. $\hat{V}_2$ is Biased for V_2

$$V_2^* = \hat{V}_2 + \hat{V}_2 [C\hat{V}_1C' - R\hat{V}_1C' - C\hat{V}_1R'] \hat{V}_2$$

$\hat{V}_1$ = Asy Var(θ_1) Based on $\ln L_1$

$\hat{V}_2$ = " " θ_2 $\qquad L_2 \mid \theta_1$

$$C = E\left[\frac{\partial \ln L_2}{\partial \theta_2} \cdot \frac{\partial \ln L_2}{\partial \theta_1'}\right]$$

$$R = E\left[\frac{\partial \ln L_2}{\partial \theta_2} \cdot \frac{\partial \ln L_1}{\partial \theta_1'}\right]$$

Estimate C using $\hat{C} = \frac{1}{T} \sum \frac{\partial \ln f_{i2}}{\partial \hat{\theta}_2} \cdot \frac{\partial \ln f_{i2}}{\partial \theta_1'}$

R use $\hat{R} = \frac{1}{T} \sum \frac{\partial \ln f_{i2}}{\partial \hat{\theta}_2} \cdot \frac{\partial \ln f_{i1}}{\partial \hat{\theta}_1'}$

Heckman (Sample Selectivity) cont.

$$y_t^o = X_t \beta + u_t \qquad \text{"regression"}$$

$$z_t^o = W_t \gamma + v_t \qquad \text{"selection"}$$

$$\begin{matrix} u_t \\ v_t \end{matrix} \sim N\left(0, \begin{matrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{matrix} \right)$$

X_t, W_t known, exogenous (or pred) variables.

$$y_t = \begin{cases} y_t^o & \text{if } z_t^o > 0 \\ \text{unobserved} & z_t^o \le 0 \end{cases}$$

$$z_t = \begin{cases} 1 & z_t^o > 0 \\ 0 & \text{otherwise.} \end{cases}$$

(log)
Likelihood Function

Two kinds of obs.

(1) $z_t = 0$

(2) $z_t = 1$ in this case $f(y_t^* \mid z_t = 1)$

Add the log probabilities of each case

$$\sum_{z_t=0} \ln \Pr(z_t = 0) + \sum_{z_t=1} \ln\left[\Pr(z_t=1) f(y_t^* \mid z_t = 1)\right]$$

$$P(A,B) = P(A|B)\,P(B)$$
$$= P(B|A)\,P(A)$$

$$\therefore \ln\left[\Pr(z_t=1)\, f(y_t^\circ \mid z_t=1)\right]$$
$$= \ln\left[\Pr(z_t=1 \mid y_t^\circ)\cdot f(y_t^\circ)\right]$$

Std result MVN.

$$\begin{matrix} X_1 \\ X_2 \end{matrix} \sim N\left(\begin{matrix} \mu_1 \\ \mu_2 \end{matrix}, \begin{matrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{matrix} \right)$$

$$X_1 \sim N(\mu_1, \Sigma_{11})$$
$$X_2 \sim N(\mu_2, \Sigma_{22})$$
$$X_1 | X_2 \sim N\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2),\ \Sigma_{1|2}\right)$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

In our case

$$\mu_2 = W_t r$$
$$\Sigma_{12} = \Sigma_{21} = \rho\sigma \quad \Sigma_{22} = \sigma^2$$

and regression becomes

$$z_t^\circ = W_t\gamma + \rho\sigma/\sigma^2(y_t^\circ - X_t\beta) + \varepsilon_t$$

$$\varepsilon_t \sim N\left(0,\ 1 - \rho\sigma\frac{1}{\sigma^2}\rho\sigma\right)$$

$$\varepsilon_t \sim N(0,\ 1-\rho^2)$$

But $z_t^\circ = \begin{cases} 1 \\ 0 \end{cases}$ and this becomes modeled as Probit.

$$\Pr(z_t = 1 \mid y_t^\circ) = \Phi\left(\frac{W_t\gamma + \frac{\rho}{\sigma}(y_t - X_t\beta)}{(1-\rho^2)^{1/2}}\right)$$

since when $z_t = 1$ $y_t^\circ = y_t$

Then log-likelihood becomes

$$\sum_{z_t=0} \log \Phi(-W_t\gamma) + \sum_{z_t=1} \log\left[\frac{1}{\sigma}\phi\left(\frac{y_t - X_t\beta}{\sigma}\right)\right]$$

$$+ \sum_{z_t=1} \log \Phi\left(\frac{W_t\gamma + \rho(y_t - X_t\beta)/\sigma}{(1-\rho^2)^{1/2}}\right)$$

Consists of

(1) Probit

(2) Normal linear Reg

(3) another term

If $\rho = 0$ Then Term $\underline{\underline{3}}$ drops out! And you can estimate each Part separately.

---

2 step Based on

$$y_t = X_t \beta + \rho \sigma v_t + e_t$$

$u_t = \epsilon_t + \sigma\rho v_t$

Since u_t & v_t have correlation $\sigma\rho$

given y_t observed

substitution yields

$$y_t = X_t\beta + \rho\sigma v_t + \epsilon_t$$

$$E(v_t \mid z_t = 1) = E(v_t \mid v_t > -w_t\gamma)$$

$$= \frac{\phi(w_t\gamma)}{\Phi(w_t\gamma)}$$

Replace v_t with conditional mean

$$y_t = X_t\beta + \rho\sigma \frac{\phi(w_t\gamma)}{\Phi(w_t\gamma)} + \epsilon_t$$

Est. γ with Probit

$$= X_t\beta + \lambda \frac{\phi(w_t\hat{\gamma})}{\Phi(w_t\hat{\gamma})} + \epsilon_t$$

$$\lambda = \rho\sigma$$