16. Tobit and SelectionA. Colin Cameron Pravin K. TrivediCopyright 2006

These slides were prepared in 1999. They cover material similar to Sections 16.2-16.3 and 16.5 of our subsequent book *Microeconometrics: Methods and Applications*, Cambridge University Press, 2005.

INTRODUCTION

- Data on y is censored if for part of the range of y we observe only that y is in that range, rather than observing the exact value of y.
 e.g. income is top-coded at \$75,000 per year.
- Data on *y* is truncated if for part of the range of *y* we do not observe *y* at all.

e.g. people with income above \$75,000 per year are excluded from the sample.

- Meaningful policy analysis requires extrapolation from the restricted sample to the population as a whole.
- But running regressions on censored or truncated data, without controlling for censoring or truncation, leads to inconsistent parameter estimates.

• We focus on the normal, with censoring or truncation at zero.

e.g. annual hours worked, and annual expenditure on automobiles.

• The class of models presented in this chapter is called limited dependent variable models or latent variable models. Econometricians also use the terminology tobit models or generalized tobit models.

- Censoring can arise for distributions other than the normal.
- For e.g. count data treatment is similar to here except different distributions.
- For duration data, e.g. the length of a spell of unemployment, a separate treatment of censoring is given there due to different censoring mechanism (random) to that considered here.

OUTLINE

- Tobit model: MLE, NLS and Heckman 2-step.
- Sample selectivity model, a generalization of Tobit.
- Semiparametric estimation.
- Structural economic models for censored choice.
- Simultaneous equation models.

TOBIT MODEL

• Interest lies in a latent dependent variable y^*

$$y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon.$$

- This variable is only partially observed.
- In censored regression we observe

 $y = \begin{cases} y^* \text{ if } y^* > 0\\ 0 \text{ if } y^* \le 0. \end{cases}$

• In truncated regression we observe

$$y = y^* \quad \text{if } y^* > 0.$$

7

- The standard estimators require stochastic assumptions about the distribution of ε and hence y^* .
- The Tobit model assumes normality:

$$\varepsilon \sim \mathsf{N}[0, \sigma^2] \Rightarrow y^* \sim \mathsf{N}[\mathbf{x}'\boldsymbol{\beta}, \sigma^2].$$

SIMULATION EXAMPLE

- Linear-log relationship between
 - -h: annual hours worked, and
 - -w: hourly wage.
- Data on desired hours of work, h^* , generated by

 $h_i^* = -2500 + 1000 \ln w_i + \varepsilon_i, \quad i = 1, ..., 250,$ $\varepsilon_i \sim \mathsf{N}[0, 1000^2],$

 $\ln w_i = 1.51, 1.52, \dots, 4.00 \Rightarrow w_i \simeq 4.5, \dots, 55.$

The wage elasticity equals 1000/h*.
e.g. 0.5 for full-time work (2000 hours).

- People work if $h^* > 0 \Rightarrow about 40\%$ do not work.
- We consider three different OLS regressions
 - Uncensored sample: regress h* on ln w.
 (In practice such data are not observed)
 - Censored sample: regress $h = \max(0, h^*)$ on $\ln w$.
 - Truncated sample: regress h^* on $\ln w$, where only observations with $h^* > 0$ are included.
- Results are presented in Table 10.1.
- Clearly 2. and 3. are inconsistent.

Variable		Sample	
	Uncensored	Censored	Truncated
ONE	-2636	-982	-382
	(256)	(174)	(297)
lnw	1043	587	477
	(90)	(61)	(95)
\mathbf{R}^2	.35	.27	.15
Observations	250	250	148

- For the censored sample:
 - Negative values of h^* have been increased to zero.
 - This increases the intercept and flattens regression line.
- For the truncated sample:
 - Observations with $\varepsilon < 0$ dropped more than those with $\varepsilon > 0$, since $\varepsilon < 0$ more likely to lead to $h^* < 0$.
 - The mean of the error is shifted up for low h.
 - This increases the intercept and flattens regression line.

• For censored and truncated data, linear regression is inappropriate.

TRUNCATED CASE: CONDITIONAL MEAN

- For truncated data we observed y only when y > 0.
- Then the truncated mean is E[y|y > 0] $= E [\mathbf{x}'\boldsymbol{\beta} + \varepsilon | \mathbf{x}'\boldsymbol{\beta} + \varepsilon > 0] \qquad \text{as } y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$ $= \mathbf{x}'\boldsymbol{\beta} + E [\varepsilon|\varepsilon > -\mathbf{x}'\boldsymbol{\beta}] \qquad \text{as } \mathbf{x} \text{ and } \varepsilon \text{ independent}$ $= \mathbf{x}'\boldsymbol{\beta} + \sigma E [(\varepsilon/\sigma)|\varepsilon/\sigma > -\mathbf{x}'(\boldsymbol{\beta}/\sigma)] \text{ transform to } \varepsilon/\sigma \sim N[0, 1]$ $= \mathbf{x}'\boldsymbol{\beta} + \sigma\lambda(\mathbf{x}'\boldsymbol{\beta}/\sigma) \qquad \text{key result for } N[0, 1].$
- $\lambda(z) = \phi(z)/\Phi(z)$ is called the inverse Mills ratio.
- The regression function is nonlinear.

• For consistent estimates use NLS or MLE, not OLS.

ASIDE: INVERSE MILLS RATIO

- Consider $z \sim N[0, 1]$, with density $\phi(z)$ and cdf $\Phi(z)$.
- The conditional density of z|z > c is $\phi(z)/(1 \Phi(c))$.
- The truncated conditional mean is $E[z|z > c] = \int_{c}^{\infty} z \left(\frac{\phi(z)}{1 - \Phi(c)} \right) dz$ $= \int_{c}^{\infty} z \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2) dz / (1 - \Phi(c))$ $= \left[-\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2) \right]_{c}^{\infty} / (1 - \Phi(c))$ $= \frac{\phi(c)}{1 - \Phi(c)} = \frac{\phi(-c)}{\Phi(-c)} = \lambda(c).$

CENSORED CASE: CONDITIONAL MEAN

- For censored data we observe y = 0 if $y^* < 0$ and $y = y^*$ otherwise.
- The censored sample mean is

$$\begin{split} \mathsf{E}[y] &= \mathsf{E}_{y^*}[\mathsf{E}[y|y^*]] \\ &= \Pr[y^* \le 0] \times 0 + \Pr[y^* > 0] \times \mathsf{E}[y^*|y^* > 0] \\ &= \Phi(\mathbf{x}'\boldsymbol{\beta}) \left\{ \mathbf{x}'\boldsymbol{\beta} + \sigma \frac{\phi(\mathbf{x}'\boldsymbol{\beta}/\sigma)}{\Phi(\mathbf{x}'\boldsymbol{\beta}/\sigma)} \right\} \\ &= \Phi(\mathbf{x}'\boldsymbol{\beta})\mathbf{x}'\boldsymbol{\beta} + \sigma \phi(\mathbf{x}'\boldsymbol{\beta}/\sigma), \end{split}$$

which is again nonlinear.

• For consistent estimates use NLS or MLE, not OLS.

MLE FOR CENSORED DATA

- Let y^* have density $f^*(y^*)$ and c.d.f. $F(y^*)$.
- Consider censored $y = \max(y^*, 0)$.
- The density for *y* is

$$- y > 0$$
: $y = y^*$ so $f(y) = f^*(y)$.

-
$$y = 0$$
: $y^* \le 0$ so $f(0) = \Pr[y^* \le 0] = F^*(0)$.

• Define indicator

$$d = \begin{cases} 1 \text{ if } y > 0\\ 0 \text{ if } y = 0. \end{cases}$$

• The density is then

 \boldsymbol{n}

$$f(y) = f^*(y)^d \times F^*(0)^{1-d}.$$

• The log-likelihood function is

$$\ln \mathsf{L} = \sum_{i=1}^{n} \left\{ d_i \ln f^*(y_i^*, \mathbf{x}_i, \boldsymbol{\theta}) + (1 - d_i) \ln F^*(0, \mathbf{x}_i, \boldsymbol{\theta}) \right\}$$

• This is a mixture of discrete and continuous densities.

CENSORED MLE FOR NORMAL

- For normal regression for notational convenience we transform from f(y) the N[x'β, σ²] density to φ(z) the N[0, 1] density.
- For y > 0

$$f(y) = f^*(y)$$

= $\left(1/\sqrt{2\pi\sigma^2}\right) \times \exp\left(-(y - \mathbf{x}'\boldsymbol{\beta})^2/2\sigma^2\right)$
= $\frac{1}{\sigma} \phi\left((y - \mathbf{x}'\boldsymbol{\beta})/\sigma\right)$,
where $\phi(z) = \left(1/\sqrt{2\pi}\right) \exp\left(-z^2/2\right)$ is N[0, 1] density.

• For
$$y = 0$$

$$f(0) = \Pr[y = 0]$$

= $\Pr[y^* \le 0]$
= $\Pr[\mathbf{x}'\boldsymbol{\beta} + \varepsilon \le 0]$
= $\Pr[\varepsilon/\sigma \le -\mathbf{x}'\boldsymbol{\beta}/\sigma]$
= $\Phi(-\mathbf{x}'\boldsymbol{\beta}/\sigma)$,

where $\Phi(z)$ is N[0, 1] c.d.f.

• Thus

$$f(y) = \left[\frac{1}{\sigma} \phi \left((y - \mathbf{x}' \boldsymbol{\beta}) / \boldsymbol{\sigma} \right) \right]^d \times \left[\Phi(-\mathbf{x}' \boldsymbol{\beta} / \boldsymbol{\sigma}) \right]^{1-d}.$$

• The log-likelihood function

$$\ln L(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \sum_{i=1}^n \left\{ d_i \ln \frac{1}{\sigma} \phi \left(\frac{y_i - \mathbf{x}'_i \boldsymbol{\beta}}{\sigma} \right) + (1 - d_i) \ln \Phi \left(-\mathbf{x}'_i \boldsymbol{\beta} / \sigma \right) \right\}$$

ASYMPTOTIC THEORY

- Tobin (1958) proposed the Tobit MLE.
- He asserted that usual ML theory applied, despite the strange continuous/discrete hybrid density.
- Amemiya (1973) provided a formal proof that usual ML theory applies, with appendix that detailed the extremum estimation approach that is now standard in econometrics.

• If density is correctly specified the usual ML theory yields after some algebra

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}}_{\mathsf{ML}} \\ \widehat{\sigma}_{\mathsf{ML}}^2 \end{bmatrix} \stackrel{a}{\sim} \mathsf{N} \begin{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \sigma^2 \end{bmatrix}, \begin{bmatrix} \sum a_i \mathbf{x}_i \mathbf{x}'_i \sum b_i \mathbf{x}'_i \\ \sum b_i \mathbf{x}'_i \sum c_i \end{bmatrix}^{-1} \end{bmatrix},$$

where defining $\boldsymbol{\alpha} = \boldsymbol{\beta}/\sigma$, $\phi_i = \phi(\mathbf{x}'_i \boldsymbol{\beta})$ and $\Phi_i = \Phi(\mathbf{x}'_i \boldsymbol{\beta})$,
 $a_i = -\frac{1}{\sigma^2} \mathbf{x}'_i \boldsymbol{\alpha} \times \phi_i + \frac{\phi_i^2}{1 - \Phi_i} - \Phi_i$
 $b_i = -\frac{1}{\sigma^2} (\mathbf{x}'_i \boldsymbol{\alpha})^2 \times \phi_i + \phi_i - \frac{\mathbf{x}'_i \boldsymbol{\alpha} \times \phi_i^2}{1 - \Phi_i}$
 $c_i = -\frac{1}{\sigma^2} (\mathbf{x}'_i \boldsymbol{\alpha})^3 \times \phi_i + \mathbf{x}'_i \boldsymbol{\alpha} \times \phi_i - \frac{\mathbf{x}'_i \boldsymbol{\alpha} \times \phi_i^2}{1 - \Phi_i} 2\Phi_i.$

MLE FOR TRUNCATED DATA

- For truncated data only $y^* > 0$ observed.
- The conditional density is

$$\begin{aligned} f^*(y^*|y^* > 0) &= f^*(y^*) / \Pr[y^*|y^* > 0] \\ &= f^*(y^*) / F^*(0). \end{aligned}$$

• The log-likelihood is then

$$\ln \mathsf{L} = \sum_{i=1}^{n} \left\{ \ln f^*(y_i^*, \mathbf{x}_i, \boldsymbol{\theta}) - \ln F^*(0, \mathbf{x}_i, \boldsymbol{\theta}) \right\}.$$

• For the normal this leads to $\binom{n}{n}$

$$\ln \mathsf{L}(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \sum_{i=1}^n \left\{ \ln \frac{1}{\sigma} \phi \left(\frac{y_i - \mathbf{x}'_i \boldsymbol{\beta}}{\sigma} \right) - \ln \frac{1}{\sigma} \Phi \left(\frac{\mathbf{x}'_i \boldsymbol{\beta}}{\sigma} \right) \right\}.$$

NLS

- Estimate using the correct censored or truncated mean.
- Recall the censored and truncated means

 $\begin{aligned} \mathsf{E}[y|y > 0] &= \mathbf{x}' \boldsymbol{\beta} + \sigma \lambda(\mathbf{x}' \boldsymbol{\beta} / \boldsymbol{\sigma}) \\ \mathsf{E}[y] &= \mathbf{x}' \boldsymbol{\beta} \times \Phi(\mathbf{x}' \boldsymbol{\beta} / \boldsymbol{\sigma}) + \sigma \phi(\mathbf{x}' \boldsymbol{\beta} / \boldsymbol{\sigma}), \quad y > 0 \end{aligned}$

- Do NLS on these, where also control for heteroskedasticity as V[y|y > 0] ≠ σ² and V[y] ≠ σ².
- Consistency requires the nonlinear functions E[y|y > 0] or E[y] are correctly specified.

- This requires strong distributional assumptions on underlying *y**.
- Any departure leads to different conditional mean functions and hence inconsistency of the NLS estimators.
 e.g. a heteroskedastic error rather than a homoskedastic error.
- Similarly for the MLE.
- This lack of robustness has led to much research.

HECKMAN TWO-STEP ESTIMATOR

- Heckman (1976, 1979) proposed estimation of the censored normal regression model by a 2-step method rather than NLS.
- Recall from that for positive y

 $\mathsf{E}[y|y>0] = \mathbf{x}'\boldsymbol{\beta} + \sigma\lambda \left(-\mathbf{x}'\boldsymbol{\beta}/\sigma\right),\,$

where $\lambda(z) = \phi(z)/\Phi(z)$ is the inverse Mills ratio.

- Heckman noted that inconsistency of OLS of y on x is due to omission of the regressor λ(-x'β/σ).
- He proposed including $\widehat{\lambda}(-\mathbf{x'}\boldsymbol{\beta}/\sigma)$ as a regressor.

- The Heckman's two-step procedure is:
 - Using censored data, estimate probit model for whether $y_i > 0$ or $y_i < 0$ with regressors \mathbf{x}_i . That is, estimate α in

 $\Pr[y_i^* > 0] = \Phi(-\mathbf{x}_i' \boldsymbol{\alpha}), \text{ where } \boldsymbol{\alpha} = \boldsymbol{\beta}/\sigma.$

Calculate the inverse Mills ratio $\lambda(\mathbf{x}'_i \widehat{\boldsymbol{\alpha}}) = \phi(\mathbf{x}'_i \widehat{\boldsymbol{\alpha}}) / \Phi(\mathbf{x}'_i \widehat{\boldsymbol{\alpha}})$.

– Using truncated data, estimate β and σ in OLS regression

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \sigma \lambda(\mathbf{x}'_i \widehat{\boldsymbol{\alpha}}) + v_i, \qquad y_i > 0.$$

- Advantages:
 - Consistent estimates using only probit and OLS.
 - Generalizes to permit weaker assumptions.
- Disadvantages:
 - Usual OLS reported standard errors are incorrect.
 - Formulae for correct standard errors take account of two complications in the OLS regression:
 - 1. Even with α known the error is heteroskedastic.
 - -2. Two-step estimator with α replaced by an estimate.
 - These corrections are complex.

DISCUSSION OF HECKMAN 2-STEP

- Aside: Note that the first-step probit only estimates β up to scale. Normally in probit we would have already normalized σ = 1. For Tobit the error variance is instead σ² and probit gives estimates of α = β/σ.
- Variations to the second step use censored not truncated regression and allow for heteroskedasticity.
- For simplest Tobit model there is little advantage to using Heckman two-step rather than NLS or the MLE.
- Advantage is in extension to more general models.

COEFFICIENT INTERPRETATION

- Interested in how the conditional mean of dependent variable changes as the regressors change.
- This varies according to whether we consider the uncensored mean, censored mean or truncated mean.
- Thus for hours worked consider effect of a change in a regressor on
 - desired hours of work,
 - actual hours of work for workers and nonworkers
 - actual hours of work for workers.

- For the standard tobit model we obtain
- For uncensored mean

$$E[y^*|\mathbf{x}] = \mathbf{x}'\boldsymbol{\beta}$$
$$\partial E[y^*|\mathbf{x}]/\partial \mathbf{x} = \boldsymbol{\beta}$$
For censored mean $(y = \max(0, y^*))$
$$E[y|\mathbf{x}] = \Phi(\mathbf{x}'\boldsymbol{\alpha})\{\mathbf{x}'\boldsymbol{\beta} + \sigma\lambda(\mathbf{x}'\boldsymbol{\alpha})\}$$

$$\partial \mathsf{E}[y|\mathbf{x}]/\partial \mathbf{x} = \Phi(\mathbf{x}'\boldsymbol{\alpha})\boldsymbol{\beta}$$

• For truncated mean(only y > 0)

$$E[y|\mathbf{x}] = \mathbf{x}'\boldsymbol{\beta} + \sigma\lambda(\mathbf{x}'\boldsymbol{\alpha})$$
$$\partial E[y|\mathbf{x}]/\partial \mathbf{x} = \{1 - (\mathbf{x}'\boldsymbol{\alpha})\lambda(\mathbf{x}'\boldsymbol{\alpha}) - \lambda(\mathbf{x}'\boldsymbol{\alpha})^2\}\boldsymbol{\beta}$$

where we use $\alpha = \beta / \sigma$, $\lambda(z) = \phi(z) / \Phi(z)$, $\partial \Phi(z) / \partial z = \phi(z)$ and $\partial \phi(z) / \partial z = -z\phi(z)$.

The censored mean expression is obtained after some manipulation. It can be decomposed into two effects (one for y = 0 and one for y > 0). See McDonald and Moffitt (1980).

SAMPLE SELECTIVITY MODEL

- Most common generalization of the standard tobit model is the sample selection or self-selection model.
- This is a two-part model
 - 1. A latent variable y_1^* that determines whether or not the process of interest is fully observed.
 - 2. A latent variable y_2^* that is of intrinsic interst.
- Classic example is labor supply
 - $-y_1^*$ determines whether or not to work and
 - y_2^* determines hours of work.

• Complication arises as the unobserved components of these two processes are correlated (after controlling for regressors).

• The two latent variables are determined as follows

$$y_1^* = \mathbf{x}_1' \boldsymbol{\beta}_1 + \varepsilon_1$$

$$y_2^* = \mathbf{x}_2' \boldsymbol{\beta}_2 + \varepsilon_2,$$

- Neither y_1^* nor y_2^* are completely observed.
- Instead we observe whether y_1^* is positive or negative $y_1 = \begin{cases} 1 \text{ if } y_1^* > 0 \\ 0 \text{ if } y_1^* \le 0. \end{cases}$ and only positive values of y_2^* $y_2 = \begin{cases} y_2^* \text{ if } y_1^* > 0 \\ 0 \text{ if } y_1^* \le 0. \end{cases}$

- The error terms are usually specified to be joint normal $\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \sim \mathsf{N} \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \end{bmatrix}.$
- For identification there mast be at least one variable included in x_2 that is not included in x_1 , or that $\sigma_{12} = 0$.
- Aside: In labor supply extra complication that even if we observe individuals with 0 hours, for these nonworkers we typically are missing data on the offered wage, a key explanatory variable. This complication can be handled by adding a third equation for the offered wage. See e.g. Mroz (1987).

MLE IN SAMPLE SELECTIVITY MODEL

- The MLE maximizes the log-likelihood function.
- This is based on the joint density $f(y_1, y_2) = f(y_2 | y_1) f(y_1)$.
 - The density $f(y_2 | y_1 = 1) = f^*(y_2 | y_1^* > 0)$ by definition. Thus $f(1, y_2) = f^*(y_2 | y_1^* > 0) \times \Pr[y_1^* > 0]$
 - The density $f(y_2|y_1 = 0)$ places probability 1 on the value $y_2 = 0$ and probability 0 on any other value since y_2 always equals 0 when $y_1 = 1$. Thus $f(0, y_1) = f(y_1|y_2 = 0) f(y_1 = 0) = \Pr[y_1^* < 0]$

Thus $f(0, y_{21}) = f(y_2 | y_1 = 0) f(y_1 = 0) = \Pr[y_1^* \le 0].$

• The likelihood function is

- $\mathsf{L} = \prod_{i=1} \left\{ \Pr[y_{1i}^* \le 0] \right\}^{1-y_{1i}} \times \left\{ f(y_{2i} \mid y_{1i}^* > 0) \times \Pr[y_{1i}^* > 0] \right\}^{y_{1i}},$
- When the bivariate density is normal, the conditional density in the second term is univariate normal and the problem is not too intractable.

HECKMAN 2-STEP: JONT NORMALITY

- Most studies use Heckman's method rather than ML.
- If the errors $(\varepsilon_1, \varepsilon_2)$ in are joint normal then

$$\varepsilon_2 = \frac{\sigma_{12}}{\sigma_1^2} \varepsilon_1 + v,$$

where v is independent of ε_1 .

• Aside: This follows from the more general result that for

$$\begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \sim \mathsf{N} \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \end{bmatrix}$$

the conditional distribution is
$$\mathbf{z}_2 | \mathbf{z}_1 \sim \mathsf{N} \begin{bmatrix} \boldsymbol{\mu}_2 + \Sigma_{21} \Sigma_{11}^{-1} (\mathbf{z}_1 - \boldsymbol{\mu}_1), \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \end{bmatrix}.$$

Thus $\mathbf{z}_2 = \boldsymbol{\mu}_1 + \Sigma_{21} \Sigma_{11}^{-1} (\mathbf{z}_1 - \boldsymbol{\mu}_1)$

• Thus $\mathbf{z}_2 = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{z}_1 - \boldsymbol{\mu}_1)$ plus a zero mean normal error independent of \mathbf{z}_1 .

• Then

$$\begin{split} \mathsf{E}[y_2|y_1^* > 0] &= \mathsf{E}[\mathbf{x}_2'\boldsymbol{\beta}_2 + \varepsilon_2|\mathbf{x}_1'\boldsymbol{\beta}_1 + \varepsilon_1 > 0] \\ &= \mathbf{x}_2'\boldsymbol{\beta}_2 + \mathsf{E}[\varepsilon_2|\varepsilon_1 > -\mathbf{x}_1'\boldsymbol{\beta}_1] \\ &= \mathbf{x}_2'\boldsymbol{\beta}_2 + \mathsf{E}\left[(\sigma_{12}/\sigma_1^2) \times \varepsilon_1 + v|\varepsilon_1 > -\mathbf{x}_1'\boldsymbol{\beta}_1\right] \\ &= \mathbf{x}_2'\boldsymbol{\beta}_2 + (\sigma_{12}/\sigma_1^2) \times \mathsf{E}[\varepsilon_1|\varepsilon_1 > -\mathbf{x}_1'\boldsymbol{\beta}_1] \\ &= \mathbf{x}_2'\boldsymbol{\beta}_2 + \frac{\sigma_{12}}{\sigma_1}\mathsf{E}\left[\frac{\varepsilon_1}{\sigma_1}|\frac{\varepsilon_1}{\sigma_1} > \frac{\mathbf{x}_1'\boldsymbol{\beta}_1}{\sigma_1}\right], \end{split}$$

where the third equality uses earlier result.

- This leads to the key result that $E[y_2|y_1^* > 0] = \mathbf{x}'_2 \boldsymbol{\beta}_2 + \frac{\sigma_{12}}{\sigma_1} \times \lambda \left(\mathbf{x}'_{1i} \boldsymbol{\beta}_1 / \sigma_1 \right),$ where $\lambda(c) = \phi(c) / \Phi(c)$ using earlier result.
- Clearly OLS of y₂ on x₂ will lead to an inconsistent estimate of β since the regressor λ(x'₁β₁/σ₁), is omitted from the equation.
- Heckman's solution is to obtain an estimate of this omitted term, and include it in the OLS regression.

- The Heckman's two-step procedure is:
 - Using censored data, estimate probit model for whether $y_{1i} > 0$ or $y_{1i} < 0$ with regressors \mathbf{x}_i . That is, estimate α_1 in

 $\Pr[y_{1i}^* > 0] = \Phi(-\mathbf{x}_{1i}'\boldsymbol{\alpha}_1), \quad \text{where } \boldsymbol{\alpha}_1 = \boldsymbol{\beta}_1/\sigma_1.$

Calculate the inverse Mills ratio $\lambda(\mathbf{x}'_{1i}\widehat{\boldsymbol{\alpha}}_1) = \phi(\mathbf{x}'_{1i}\widehat{\boldsymbol{\alpha}}_1)/\Phi(\mathbf{x}'_{1i}\widehat{\boldsymbol{\alpha}}_1)$

– Using truncated data on y_2 , estimate β_2 and σ in OLS regression

$$y_{2i} = \mathbf{x}'_{2i}\boldsymbol{\beta}_2 + (\sigma_{12}/\sigma_1)\lambda(\mathbf{x}'_{1i}\widehat{\boldsymbol{\alpha}}_1) + v_i, \qquad y_{2i} > 0.$$

• As in the classic Tobit model the resulting estimators

of β_2 are consistent, inefficient and it is cumbersome to construct the variance covariance matrix of OLS.

HECKMAN 2-STEP: WEAKER ASSUMPTIONS

- The Heckman two-step method relies on weaker distributional assumptions than the MLE.
- The MLE requires joint normality of ε_1 and ε_2 .
- The Heckman 2-step estimator requires the weaker assumption that

$$\varepsilon_2 = \delta \varepsilon_1 + v,$$

where v is independent of ε_1 and ε_1 is normally distributed.

• In the case of purchase of a durable good, this says

that the error in the amount purhased equation is a multiple of the error in the purchase decision equation, plus some noise where the noise is independent of the purchase decision.

• Given this assumption we obtain

 $\mathsf{E}[y_2|y_1^*>0] = \mathbf{x}_2'\boldsymbol{\beta}_2 + \delta\mathsf{E}[\varepsilon_1|\varepsilon_1>-\mathbf{x}_1'\boldsymbol{\beta}_1].$

- Heckman's two-step method can be adapted to
 - distributions for ε_1 other than normal
 - semiparametric methods which do not impose a functional form for $E[\varepsilon_1|\varepsilon_1 > -\mathbf{x}'_1\boldsymbol{\beta}_1]$.

COEFFICIENT INTERPRETATION

- Again interested in how the conditional mean of dependent variable changes as the regressors change.
- This varies according to whether we consider the uncensored mean, censored mean or truncated mean.

• For uncensored mean

$$\mathsf{E}[y_2^*|\mathbf{x}] = \mathbf{x}_2'\boldsymbol{\beta}_2$$
$$\partial E[y_2^*|\mathbf{x}_2]/\partial \mathbf{x}_2 = \boldsymbol{\beta}_2$$

• For censored mean

 $\begin{aligned} \mathsf{E}[y_2|y_1^* > 0, \mathbf{x}] &= \mathbf{x}' \boldsymbol{\beta}_2 + (\sigma_{12}/\sigma_1^2)\lambda(\mathbf{x}' \boldsymbol{\beta}_1/\sigma_1)] \\ \partial \mathsf{E}[y_2|y_1^* > 0, \mathbf{x}]/\partial \mathbf{x} &= (\boldsymbol{\beta}_1/\sigma_1)\phi(\mathbf{x}_1' \boldsymbol{\beta}_1/\sigma_1)[\mathbf{x}' \boldsymbol{\beta}_2 + (\sigma_{12}/\sigma_1^2)\lambda(\mathbf{x}' \boldsymbol{\beta}_1/\sigma_1) \\ &+ \Phi(\mathbf{x}' \boldsymbol{\beta}_1/\sigma_1)[\boldsymbol{\beta}_2 + (\sigma_{12}/\sigma_1^2)\partial\lambda(\mathbf{x}' \boldsymbol{\beta}_1/\sigma_1)/\partial \mathbf{x}] \end{aligned}$

• For truncated mean...

$$\mathsf{E}[y_2|y_1^* > 0, \mathbf{x}] = \mathbf{x}'\boldsymbol{\beta}_2 + (\sigma_{12}/\sigma_1^2)\lambda(\mathbf{x}'\boldsymbol{\beta}_1/\sigma_1)$$
$$\partial \mathsf{E}[y_2|y_1^* > 0, \mathbf{x}]/\partial \mathbf{x} = \mathbf{x} + (\sigma_{12}/\sigma_1^2)\partial\lambda(\mathbf{x}'\boldsymbol{\beta}_1/\sigma_1)/\partial \mathbf{x}].$$

SEMIPARAMETRIC ESTIMATION

- Consistency of all the above estimators requires correct specification of the error distribution.
- Any misspecification of the error distribution leads to inconsistency e.g. failure of normality.
- So preferable to have an estimator that does not require specification of the distribution of the error.
- A number of semiparametric estimators that do not require distribution of the error distribution have been proposed.

- For the standard tobit model this has been done.
- Unfortunately, this model is often too simple and the generalized tobit model needs to be used. Then there are fewer results on semiparametric estimators.
- A recent application for the sample selectivity model is given in Newey, Powell and Walker (1990).
- This is a major area of current research by theoretical econometricians.
- These often use Heckman's two-step framework.

MIXED DISCRETE/CONTINUOUS MODELS

• Hanemann (1984) and Dubin and McFadden (1984) develop models from economic theory of utility maximization.

SIMULTANEOUS EQUATIONS TOBIT MODELS

• An example is the generalized tobit model, with latent variables

$$y_1^* = \mathbf{x}_1' \boldsymbol{\beta}_1 + \alpha_1 y_2^* + \gamma_1 y_2 + \varepsilon_1$$
$$y_2^* = \mathbf{x}_2' \boldsymbol{\beta}_2 + \alpha_2 y_1^* + \gamma_2 y_1 + \varepsilon_2$$

- This has the added complication that regressors in the first equation include y_2^* or y_2 and similarly y_1 or y_1^* in the second equation. Identification conditions will of course not permit all of these variables to be included.
- Treatment of simultaneity in these models is very difficult.

- Often very ingenious methods allow estimation using just regular probit, tobit and OLS commands.
- But getting the associated standard errors of estimators is very difficult.
- The simplest model has as right-hand side endogenous variables only the latent variables y_2^* or y_1^* .
- We can then obtain a reduced form for y_1^* and y_2^* , in exactly the same way as regular linear simultaneous equations, and do tobit estimation on this reduced form.
- Estimators for this case are given in Nelson and Olson

(1978), Amemiya (1979), and Lee (1981).

- Treatment is much more difficult when right-hand side endogenous variables are the observed variables y_2 or y_1 .
- See Heckman (1978) and Blundell and Smith (1989).
- Simultaneity in tobit (and probit) models can generally be handled, but can require a considerable degree of econometric sophistication.
- If possible, specify the models such that the simultaneity is due to the latent variables and not the observed variables.

APPLICATION: LABOR SUPPLY

- Use data of Mroz (1987) on 753 married women from the 1976 Panel Survey of Income Dynamics (PSID).
- Dependent variable HOURS is annual hours worked in previous year. For this sample there is a bunching or censoring at zero since 225 (or 43%) had zero hours.

- The regressors are a constant term and
- 1. KL6: Number of children less than six
- 2. K618: Number of children more than six
- 3. AGE: Age
- 4. ED: Education (years of schooling completed)

5. NLINCOME: annual nonlabor income of wife measured in \$10,000's.

Variable		Coeff			t-stat	
	COLS	TOLS	Tobit	COLS	TOLS	Tobit
ONE	1346	2118	1111	4.7	6.2	2.3
KL6	-504	-341	-1060	-7.9	-3.4	-8.4
K618	-82	-114	-106	-3.3	-3.7	-2.5
AGE	-20	-8	-36	-4.4	-1.4	-4.7
ED	57	-16	127	4.1	-1.0	-5.4
NLINCOME	-104	-43	-220	-1.2	-4.4	-4.5
n	753	428	753	753	428	753

- Estimates from censored OLS (COLS), truncated OLS (TOLS) and censored tobit (tobit) with associated t-ratios are presented in the table.
- As observed earlier, both censoring and truncation flatten the slope coefficients.
- The truncated regression suggests that some variables such as AGE, ED and NLINCOME may have more impact on the decision whether to work or not than an actual hours of work.

ASYMPTOTIC THEORY FOR HECKMANS'S 2-STEP METHOD

- Two different methods, which give the same result, are presented.
 - One method is specific to least squares type estimators.
 - The second method is a general method for any 2step estimator, including those for highly nonlinear models.
- We wish to estimate the parameters $\theta = (\beta', \sigma)'$ in the

equation

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \sigma \lambda(\mathbf{x}'_i \boldsymbol{\alpha}) + \eta_i,$$

where $\eta_i = y_i - \mathbf{x}'_i \boldsymbol{\beta} - \sigma \lambda(\mathbf{x}'_i \boldsymbol{\alpha})$ is heteroskedastic with variance $\sigma_{\eta i}^2$ defined in ?? The first step of the two-step procedure is to obtain an estimate $\hat{\alpha}$ of the unknown parameter $\boldsymbol{\alpha}$. The second step is to estimate by OLS the model

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \sigma \lambda(\mathbf{x}'_i \widehat{\boldsymbol{\alpha}}) + v_i,$$

where

$$v_i = \eta_i + \sigma(\lambda(\mathbf{x}'_i \widehat{\boldsymbol{\alpha}}) - \lambda(\mathbf{x}'_i \boldsymbol{\alpha})).$$

- First method of proof.
- Rewrite the second-step model as

$$y_i = \widehat{\mathbf{w}}_i' \boldsymbol{\theta} + v_i$$

where $\widehat{\mathbf{w}}_i = (\mathbf{x}_i, \lambda(\mathbf{x}'_i \widehat{\boldsymbol{\alpha}}))$, or in matrix notation $y = \widehat{\mathbf{W}} \boldsymbol{\theta} + v.$

- By the usual techniques the OLS estimator $\hat{\theta} = (\widehat{\mathbf{W}}'\widehat{\mathbf{W}})^{-1}\widehat{\mathbf{W}}'\mathbf{y}$ can be re-expressed as $\sqrt{n}(\widehat{\theta} - \theta) = (n^{-1}\widehat{\mathbf{W}}'\widehat{\mathbf{W}})^{-1}n^{-1/2}\widehat{\mathbf{W}}'\mathbf{v}.$
- Now plim $n^{-1}\widehat{\mathbf{W}'}\widehat{\mathbf{W}} = \lim n^{-1}\mathbf{W'}\mathbf{W}$, where $\mathbf{w}'_i = (\mathbf{x}'_i, \lambda(\mathbf{x}'_i \boldsymbol{\alpha}))'$.
- The hard part is to obtain the limit distribution of

 $n^{-1/2}\widehat{\mathbf{W}}\mathbf{v}$. By a first-order Taylor series expansion the error term:

$$v_i = \eta_i + \frac{\partial \lambda_i}{\partial \alpha'} (\widehat{\alpha} - \alpha)$$

which is both heteroskedastic via η_i and potentially correlated via the second term. It is obvious that v_i asymptotically has zero mean. It can be shown that the first and second terms on the right-hand are asymptotically uncorrelated, and we just consider the two terms in isolation.

• It follows that

 $n^{-1/2}\widehat{\mathbf{W}}v \xrightarrow{d} \mathsf{N}\left[0, \lim n^{-1}\mathbf{W}\mathsf{E}[\eta\eta']\mathbf{W} + \lim n^{-1}\mathbf{W}\mathbf{D}\mathbf{V}_{\alpha}\mathbf{D}\mathbf{W}\right]$

where D has i^{th} row $D_i = \partial \lambda_i / \partial \alpha$ and $\hat{\alpha}$ is asymptotically N[α , V $_{\alpha}$].

• Combining these results gives the Heckman two-step estimator

$$\widehat{\boldsymbol{\theta}} \stackrel{a}{\sim} \mathsf{N}\left(\boldsymbol{\theta}, \mathbf{V}_{\boldsymbol{\theta}}\right),$$

where V_{θ} is consistently estimated by

$$\widehat{\mathbf{V}}_{\boldsymbol{\theta}} = (\widehat{\mathbf{W}}'\widehat{\mathbf{W}})^{-1} (\widehat{\mathbf{W}}'\Sigma_{\widehat{\eta}}\widehat{\mathbf{W}} + \widehat{\mathbf{W}}'\widehat{\mathbf{D}}\mathbf{V}_{\boldsymbol{\alpha}}\widehat{\mathbf{D}}\widehat{\mathbf{W}}) (\widehat{\mathbf{W}}'\widehat{\mathbf{W}})^{-1},$$

where
$$\widehat{\mathbf{W}}'\widehat{\mathbf{W}} = \sum_{i=1}^{n} \widehat{\mathbf{w}}_{i}\widehat{\mathbf{w}}_{i}', \ \widehat{\mathbf{W}}'\widehat{\mathbf{D}} = \sum_{i=1}^{n} \widehat{\mathbf{w}}_{i}\widehat{\mathbf{d}}_{i}, \ \widehat{\mathbf{d}}_{i} = \partial \lambda_{i}(\mathbf{x}_{i}'\boldsymbol{\alpha})/\partial \boldsymbol{\alpha} \mid_{\widehat{\boldsymbol{\alpha}}}$$

and $\Sigma_{\widehat{\boldsymbol{\eta}}}$ is a diagonal matrix with i^{th} entry $\widehat{\sigma}_{\boldsymbol{\eta}_{i}}^{2}$.

- This estimate is straightforward to obtain if matrix commands are available. The hardest part can be analytically obtaining $\sigma_{\eta_i}^2 = V[\eta_i]$.
- If this is difficult we can instead use $\hat{\sigma}_i^2 = (y_i \mathbf{x}'_i \hat{\boldsymbol{\beta}} + \hat{\sigma} \lambda_i (\mathbf{x}'_i \hat{\boldsymbol{\alpha}}))$ following the approach of White (1980).

• Second method of proof is to write the normal equations for the two parts of the two-step estimator as

$$\begin{split} \sum_{i=1}^{n} g(\mathbf{x}_{i}, \boldsymbol{\alpha}) &= \mathbf{0} \\ \sum_{i=1}^{n} \left[\begin{array}{c} \mathbf{x}_{i} \\ \lambda_{i} \end{array} \right] \left(y_{i} - \mathbf{x}_{i}^{\prime} \boldsymbol{\beta} + \lambda_{i} (\mathbf{x}_{i}^{\prime} \boldsymbol{\alpha}) \boldsymbol{\sigma} \right) &= \mathbf{0}, \end{split}$$

where the first equation is the first-order conditions for α and the second equation gives OLS first-order conditions for β . • These equations can be combined as

$$\sum_{i=1}^{n} \mathbf{q}(\mathbf{x}_{i}, \gamma) = \mathbf{0},$$

where $\boldsymbol{\gamma} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'.$

• By the usual first-order Taylor series expansion

$$\widehat{\boldsymbol{\gamma}} \stackrel{a}{\sim} \mathsf{N}\left[\boldsymbol{\gamma}, \left[\sum_{i=1}^{n} \frac{\partial \mathbf{q}(\mathbf{x}_{i}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}'}\right]^{-1} \left[\sum_{i=1}^{n} \mathbf{q}(\mathbf{x}_{i}, \boldsymbol{\gamma}) \mathbf{q}(\mathbf{x}_{i}, \boldsymbol{\gamma})'\right] \left[\sum_{i=1}^{n} \frac{\partial \mathbf{q}(\mathbf{x}_{i}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\right]$$

 We are interested in the sub-component corresponding to β.

- Simplification occurs because ∂q(x_i, γ)/∂γ is block triangular because β does not appear in the first set of equations. Newey (1984) gives the simpler formulae for this case.
- Applying it to the example here will give the result given earlier. Related papers are Pagan (1986) and Murphy and Topel (1985).