CORNER SOLUTIONS: TOBIT MODELS

Econometric Analysis of Cross Section and Panel Data, 2e
MIT Press
Jeffrey M. Wooldridge

1. Introduction
2. A General Formulation
3. The Type I Tobit Model
4. Specification Issues
5. Estimating Parameters Under Weaker Assumptions
6. Endogenous Explanatory Varibles
7. Panel Data
8. Two-Limit Tobit Models

1. INTRODUCTION

- So-called "censored regression models" are applied to two very different kinds of situations: (1) data censoring, such as top coding of wealth or censoring of a duration (survival time); (2) corner solution outcomes, where the variable we would like to explain piles up at one or two corners.

- Most applications are to the latter case, where we always observe the response, but it takes on values at a corner (often zero, sometimes one, and occasionally other values). It is continuous (or roughly so) over strictly positive values.

- Examples include annual charitable contributions, annual labor supply, amount of life insurance.
- Count variables (number of crimes committed, annual medical appointments, patents awarded) are not examples because, while they take on the value zero, they are not roughly continuous over strictly postive values. (Typically, they take on small positive integer values.) We have used the Poisson MLE as an example for count data; count data models are covered in detail in Chapter 18.

- Before applying any limited dependent variable model we should ask, “Do I always observe the response variable of interest?” For variables such as charitable contributions and annual hours worked, the answer is usually “yes” because an observed zero means zero.
- In true data censoring cases, the answer is “no,” even though unobserved values are often assigned to a censoring value.

- We will cover censored data, where we have a linear model for the response variable we would like to explain but the data have been censored, in Chapter 19.
- More subtle situations arise. For example, we might think charitable contributions follows a particular population model that allows a corner at zero – such as the Type I Tobit model in Section 4 – but also have data censoring (say, top coding).

- Little confusion arises if one first specifies the model for the underlying population and separately describes the mechanism generating the data. The latter can be data censoring or where we cannot randomly sample units from the population of interest.
- Here we are interested only in specifying population models that are logically consistent with the population distribution of corner solution responses.

- When we turn to two-part or hurdle models, we also need to be aware that there is confusion between cases where we always observe the response of interest and where we do not for some subset of the population. Nonrandom sample selection from the underlying population are covered in Chapter 19.

- Consider the leading case where we want to determine factors that affect y, where $y \geq 0$ with $P(y = 0) > 0$ but y is (roughly) continuous over strictly positive values. So y is neither discrete nor continuous.
- Is there anything necessarily wrong with a linear model for $E(y|\mathbf{x})$? In fact, if we are interested in estimating partial effects of the x_j on $E(y|\mathbf{x})$, a linear model is a reasonable starting point.

- The drawbacks of a linear model when y is a corner solution are similar to those for the linear probability model: (1) Some of the fitted values can be negative; (2) the partial effects are unlikely to be constant througout the range of $\mathbf{x}$; (3) $Var(y|\mathbf{x})$ is unlikely to be constant. The second is the most important.
- Just as with other limited dependent variable, when y is a corner solution response we should view the linear model as representing the linear projection, $L(y|\mathbf{x}) = \mathbf{x}\boldsymbol{\gamma}$. Whether the the LP is a good approximation to $E(y|\mathbf{x})$ is an important question.

- Alternative functional forms might provide a better approximation. It is not crazy to use, say, $E(y|\mathbf{x}) = \exp(\mathbf{x}\boldsymbol{\beta})$ and apply nonlinear least squares (or a quasi-MLE of the type covered in Chapter 18). But the exponential functional form does not fall out of standard models for corners. We study similar possibilities under two-part models.
- Another issue with corner solution responses (which does not arise with binary responses) is that there are more features of $D(y|\mathbf{x})$ that we can be interested in, such as $E(y|\mathbf{x}, y > 0)$ or $P(y > 0|\mathbf{x})$.

2. A GENERAL FORMULATION

• Consider the case where $y \geq 0$ has a corner at zero. A useful starting point is

$$y = \max(0, \mathbf{x}\boldsymbol{\beta} + u)$$

where $\mathbf{x} = (1, x_2, \ldots, x_K)$, $\boldsymbol{\beta}$ is $K \times 1$, and u is an unobserved error with a continuous distribution.

• If the range of u is unrestricted, this setup generates a pile up at zero and then continuous strictly positive outcomes.

• Benefit of this approach: it keeps the focus on y. The question is what this model implies about $D(y|\mathbf{x})$.

• The phrase "y is censored at zero" is common, but not ideal. Censoring implies a missing data problem, but there is none here. "y has a corner at zero" is more descriptive of this situation.

• What can we say about $D(y|\mathbf{x})$ in general? Not much until we restrict $D(u|\mathbf{x})$ in some way. For example, assume

$$Med(u|\mathbf{x}) = 0,$$

which holds in common parametric models. Then, because $g(z) \equiv \max(0, z)$ is a nondecreasing function on $\mathbb{R}$, we can pass the median through:

$$Med(y|\mathbf{x}) = \max(0, \mathbf{x}\boldsymbol{\beta} + Med(u|\mathbf{x})) = \max(0, \mathbf{x}\boldsymbol{\beta}).$$

- This result is the basis for so-called *censored least absolute deviations (CLAD)* estimation of $\boldsymbol{\beta}$ – later.
- Generally, we cannot find $E(y|\mathbf{x})$ without much stronger assumptions, even if we make the standard assumption $E(u|\mathbf{x}) = 0$. However, because $g(z) = \max(0, z)$ is convex, we can use Jensen's inequality:

$$E(y|\mathbf{x}) = E[\max(0, \mathbf{x}\boldsymbol{\beta} + u)|\mathbf{x}] \geq \max(0, \mathbf{x}\boldsymbol{\beta} + E(u|\mathbf{x})) \quad = \max(0, \mathbf{x}\boldsymbol{\beta}).$$

So

$$E(y|\mathbf{x}) \geq \max(0, \mathbf{x}\boldsymbol{\beta}).$$

- Even if we could estimate $\boldsymbol{\beta}$ this inequality is not helpful for estimating partial effects on $E(y|\mathbf{x})$.

• Sometimes a latent variable formulation is helpful, where y^* is the latent variable:

$$\begin{aligned} y^* &= \mathbf{x}\boldsymbol{\beta} + u \\ y &= \max(0, y^*) \end{aligned}$$

But usually y^* is not the variable of interest. ("Desired" charitable contributions or "desired" labor supply are not what policy makers care about. It is the actual outcomes.)

• If $E(u|\mathbf{x}) = 0$, $\beta_j = \partial E(y^*|\mathbf{x})/\partial x_j$, but we are more interested in $E(y|\mathbf{x})$.

3. THE TYPE I TOBIT MODEL

- By far the most popular model for corners at zero assumes normality of the latent error:

$$\begin{aligned} y &= \max(0, \mathbf{x}\boldsymbol{\beta} + u) \\ u|\mathbf{x} &\sim Normal(0, \sigma^2) \end{aligned}$$

- So, if we write $y^* = \mathbf{x}\boldsymbol{\beta} + u$, then $D(y^*|\mathbf{x})$ follows a classical linear model.

Quantities of Interest and Partial Effects

• Under these assumptions, we have fully characterized $D(y|\mathbf{x})$. Let $w = 1[y > 0]$ be the binary indicator for whether y is positive or at the corner. Then

$$\begin{aligned} P(w = 1|\mathbf{x}) &= P(\mathbf{x}\boldsymbol{\beta} + u > 0|\mathbf{x}) = P(u/\sigma > -\mathbf{x}\boldsymbol{\beta}/\sigma|\mathbf{x}) \\ &= 1 - \Phi(-\mathbf{x}\boldsymbol{\beta}/\sigma) = \Phi(\mathbf{x}\boldsymbol{\beta}/\sigma). \end{aligned}$$

• So w given $\mathbf{x}$ follows a probit with parameter vector $\boldsymbol{\beta}/\sigma$, so we already know how to obtain partial effects on $P(y > 0|\mathbf{x})$. If x_j is continuous,

$$\frac{P(y > 0|\mathbf{x})}{\partial x_j} = (\beta_j/\sigma)\phi(\mathbf{x}\boldsymbol{\beta}/\sigma).$$

- Fact: If $z \sim Normal(0,1)$ then $E(z|z > c) = \phi(c)/[1 - \Phi(c)]$. So

$$\begin{aligned}
E(y|\mathbf{x}, y > 0) &= \mathbf{x}\boldsymbol{\beta} + E(u|u > -\mathbf{x}\boldsymbol{\beta}) \\
&= \mathbf{x}\boldsymbol{\beta} + \sigma E(u/\sigma|u/\sigma > -\mathbf{x}\boldsymbol{\beta}/\sigma) \\
&= \mathbf{x}\boldsymbol{\beta} + \sigma\left[\frac{\phi(-\mathbf{x}\boldsymbol{\beta}/\sigma)}{1 - \Phi(-\mathbf{x}\boldsymbol{\beta}/\sigma)}\right] \\
&= \mathbf{x}\boldsymbol{\beta} + \sigma\left[\frac{\phi(\mathbf{x}\boldsymbol{\beta}/\sigma)}{\Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)}\right] \\
&\equiv \mathbf{x}\boldsymbol{\beta} + \sigma\lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)
\end{aligned}$$

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

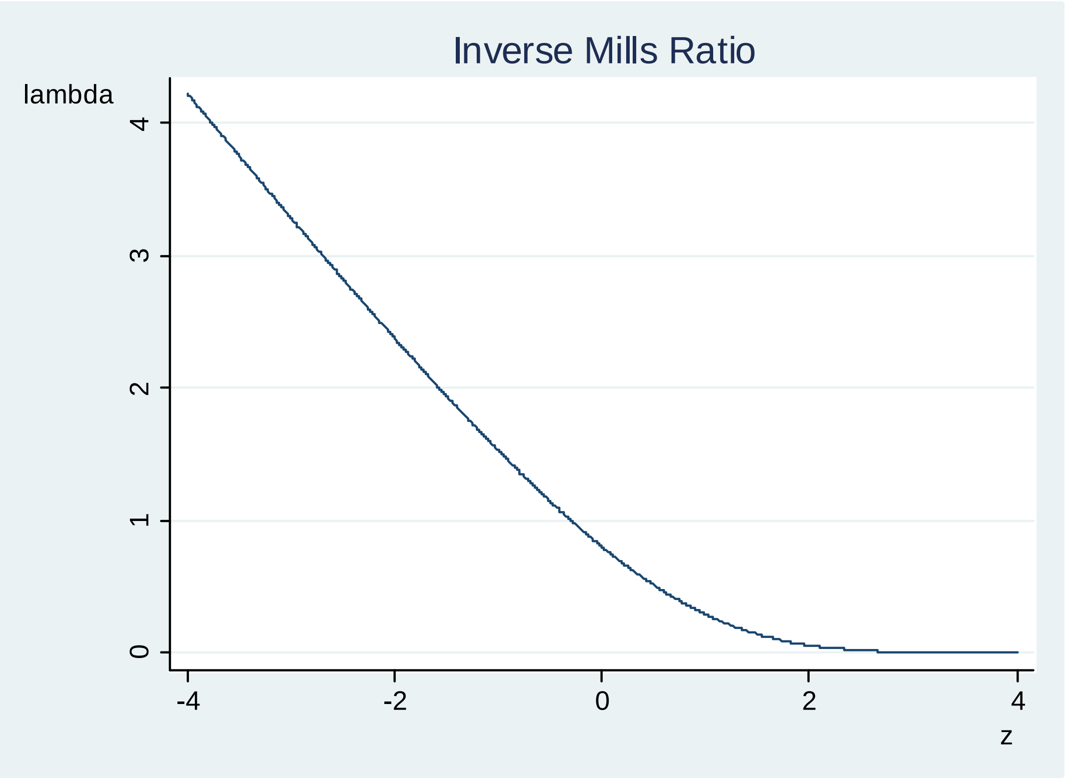
- $\lambda(\cdot)$ is strictly decreasing, and looks linear over much of its range.
- Because $\phi(z) \to 0$ and $\Phi(z) \to 1$ as $z \to \infty$, $\lim_{z\to\infty} \lambda(z) = 0$. Further,

using L'Hôspital's rule it is easy to show that

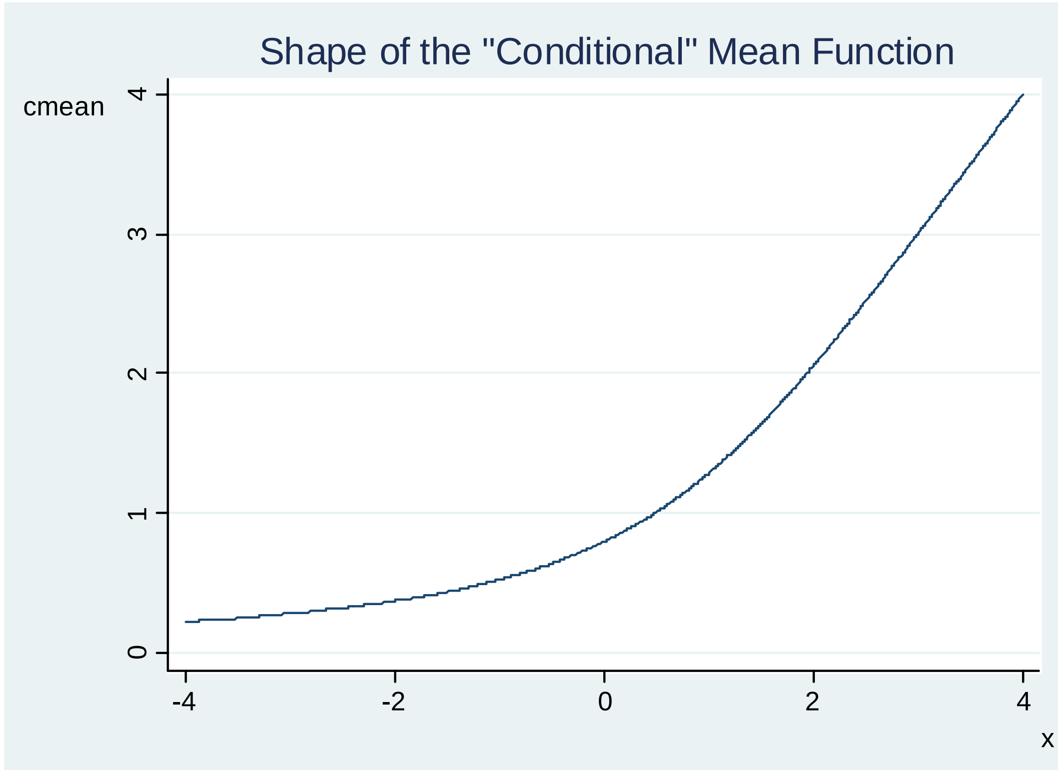
$$\lim_{z\to-\infty} \lambda(z) = \infty.$$

- The proof uses

$$\frac{d\phi}{dz}(z) = -z\phi(z) \text{ and } \frac{d\Phi}{dz}(z) = \phi(z).$$


Inverse Mills Ratio
lambda
0
1
2
3
4
-4
-2
0
2
4
z

- The IMR appears often in models with LDVs and, later, in data censoring and sample selection contexts.
- $E(y|\mathbf{x}, y > 0)$ is called the "conditional" expectation because it is conditional on $y > 0$. Of course, it is also conditional on $\mathbf{x}$.

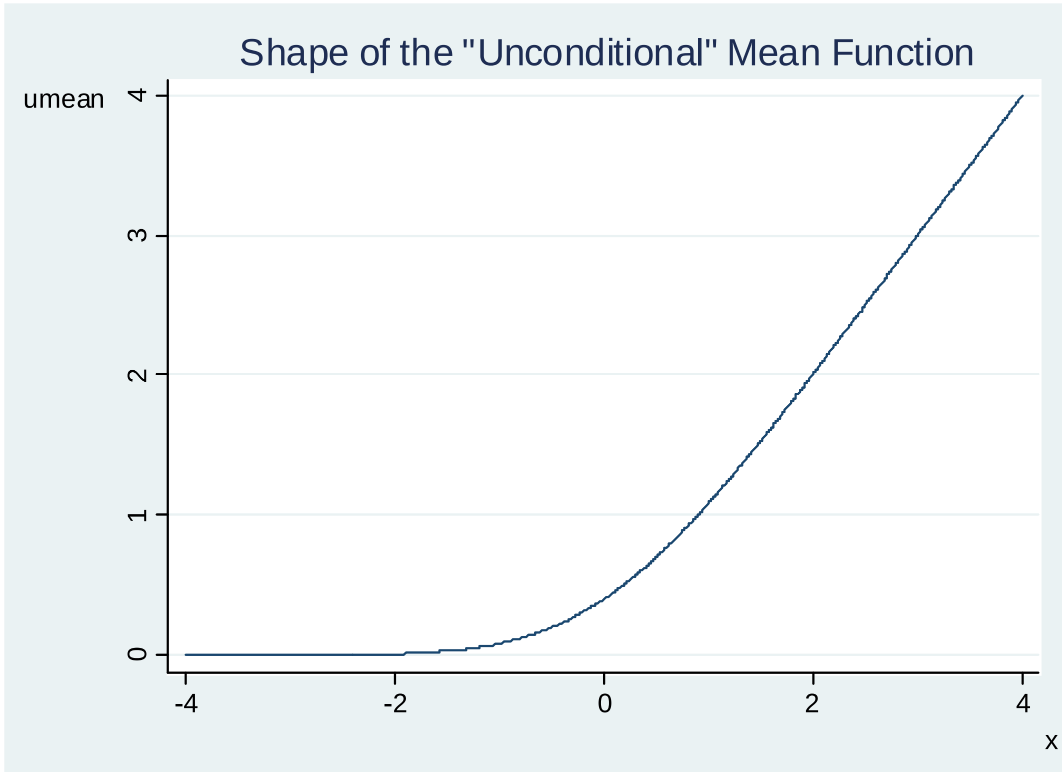


Shape of the "Conditional" Mean Function
cmean
0
1
2
3
4
-4
-2
0
2
4
x

- Now we can use

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

- Called the “unconditional” expectation (but, as always, we are conditioning on **x**).


Shape of the "Unconditional" Mean Function
umean
0
1
2
3
4
-4
-2
0
2
4
x

- Partial effects on $P(y > 0|\mathbf{x})$ already known from probit:

$$\frac{\partial P(y > 0|\mathbf{x})}{\partial x_j} = (\beta_j/\sigma)\phi(\mathbf{x}\boldsymbol{\beta}/\sigma)$$

- Partiel effects on $E(y|\mathbf{x}, y > 0)$: uses $d\lambda(c)/dc = -\lambda(c)[c + \lambda(c)]$.

$$\begin{aligned}\frac{E(y|\mathbf{x}, y > 0)}{\partial x_j} &= \beta_j - \beta_j\lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)[\mathbf{x}\boldsymbol{\beta}/\sigma + \lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)] \\ &= \beta_j\{1 - \lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)[\mathbf{x}\boldsymbol{\beta}/\sigma + \lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)]\} \\ &\equiv \beta_j\theta(\mathbf{x}\boldsymbol{\beta}/\sigma)\end{aligned}$$

- If x_j and x_h are two continuous variables, the ratio of partial effects is β_j/β_h, free of $\mathbf{x}$.

• Can show that $\theta(c) \equiv 1 - \lambda(c)[c + \lambda(c)] = Var(z|z > -c)$ where z is standard normal, so $0 < \theta(c) < 1$ for all c.

• This implies that $\lambda(c)[c + \lambda(c)] > 0$ for all c, and so the IMR, whose slope is $-\lambda(c)[c + \lambda(c)]$, is strictly decreasing.

• β_j gives direction of effect, but must be scaled down. To compare with the linear projection $L(y|\mathbf{x})$ estimated by a linear model, must scale down the Tobit estimates to make them comparable to OLS.

• For discrete changes, need to compare the expected values at two different values of covariates. Because the conditional mean function $a + \sigma\lambda(a/\sigma)$ is strictly increasing in a for any $\sigma > 0$, sign of β_j gives direction of ceteris paribus effect.

- For the unconditional expectation, a generally useful experssion is

$$\frac{E(y|\mathbf{x})}{\partial x_j} = \frac{\partial P(y > 0|\mathbf{x})}{\partial x_j} \cdot E(y|\mathbf{x}, y > 0) + P(y > 0|\mathbf{x}) \cdot \frac{E(y|\mathbf{x}, y > 0)}{\partial x_j}.$$

- Applied to the Tobit model, after some algebra, we get

$$\frac{\partial E(y|\mathbf{x})}{\partial x_j} = \Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)\beta_j = P(y > 0|\mathbf{x})\beta_j.$$

- So, to get partial effects on the unconditional mean, again the β_j are scaled by a function between 0 and 1 (and which depends on **x**).

- As $P(y > 0|\mathbf{x}) \rightarrow 1$ the β_j become close to the actual partial effect. If $P(y = 0|\mathbf{x})$ is large, the scale factor is small.
- Any comparison across models – linear, two-part, and so on – must account for any scale factors (assuming the partial effects have such a simple form).

• We can show that the expected value of $\partial E(y|\mathbf{x})/\partial x_j$ is simply $P(y > 0)\beta_j$ because a probability (conditional or unconditional) can be written as the expected value of an indicator function. Then we apply iterated expectations:

$$E[P(y > 0|\mathbf{x})] \;= E\{E(1[y > 0]|\mathbf{x})\} \;= E(1[y > 0]) \;= P(y > 0).$$

- In other words, the APE on $E(y|\mathbf{x})$ for a continuous variable x_j is just

$$APE_j = P(y > 0)\beta_j$$

We can always consistently estimate $P(w = 1) = P(y > 0)$ as the fraction of nonzero outcomes in the sample. (This is true more generally; not just for the Tobit model.)

Estimation of Parameters

• Suppose now we have a random sample $\{(\mathbf{x}_i, y_i) : i = 1, 2, \ldots, N\}$ from the population. From the expressions for $E(y|\mathbf{x}, y > 0)$ and $E(y|\mathbf{x})$, it is pretty clear that OLS regressions y_i on $\mathbf{x}_i$, using either the full sample or the sample with $y_i > 0$, does not consistently estimate $\boldsymbol{\beta}$. There are some results about estimating these parameters up to a common scale. However, when y is directly of interest, we are more interested in the regression coefficients themselves (which approximate the average partial effects).

- These days, MLE (conditional on $\mathbf{x}$) is the only way to go for Type I Tobit. The log likelihood can be parameterized to be concave.
- We need the density for y given $\mathbf{x}$. But

$$f(0|\mathbf{x}) = 1 - \Phi(\mathbf{x}\boldsymbol{\beta}/\sigma).$$

For $\mathfrak{y} > 0$, $f(\mathfrak{y}|\mathbf{x}) = f^*(\mathfrak{y}|\mathbf{x})$, where f^* is the density of the latent variable. So we have

$$f(\mathfrak{y}|\mathbf{x}) = [1 - \Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)]^{1[\mathfrak{y}=0]}[\sigma^{-1}\phi[(\mathfrak{y} - \mathbf{x}\boldsymbol{\beta})/\sigma]^{1[\mathfrak{y}>0]}$$

(where $\mathfrak{y}$ is just a dummy argument in the density).

• The log likelihood for random draw i is

$$\ell_i(\boldsymbol{\beta},\sigma) = 1[y_i = 0]\log[1 - \Phi(\mathbf{x}_i\boldsymbol{\beta}/\sigma)] \\ + 1[y_i > 0]\{\log\phi[(y_i - \mathbf{x}\boldsymbol{\beta})/\sigma] - \log(\sigma)\}$$

and, as usual, we sum across all i to get the log likelihood for the sample.

• See the text for score and Hessian. It is a well-behaved maximization problem, and software does this routinely without computational difficulty.

- In Stata:

```
tobit y x1 x2 ... xK, ll(0)
```

where the qualifier “ll(0)” means that the lower limit is at zero. (One can specify a different lower limit and also an upper limit, but this happens more often for data censoring rather than corner solution responses.)

- If we use a “robust” option in Stata (and other packages) then – just as with probit, logit, multinomial logit, and the ordered models – then we are admitting that the model is misspecified. If the Tobit model is correct for $D(y|\mathbf{x})$ we do not need a sandwich form of the asymptotic variance.

- If we use a sandwich form, we can perform inference on the parameters in the misspecified model. (More precisely, on the plims of the quasi-MLEs.)
- Another possible justification of a sandwich variance matrix estimator is that it might produce better standard errors than the estimated inverse of the expected Hessian. But this does not appear to be true for Tobit models.

Reporting the Results

• We obtain estimates $\hat{\beta}_j$ along with standard errors. We also obtain $\hat{\sigma}$ and its standard error. The latter parameter shows up in partial effects, so it is not "ancillary."

• We can report estimated PEAs on, say, $E(y|\mathbf{x})$, as

$$\Phi(\bar{\mathbf{x}}\hat{\boldsymbol{\beta}}/\hat{\sigma})\hat{\beta}_j$$

for continuous variables. (Unfortunately, the "mfx" command in Stata simply reports the $\hat{\beta}_j$.) Or, plug in other interesting values.

- The estimated APEs are

$$\left[N^{-1} \sum_{i=1}^{N} \Phi(\mathbf{x}_i \hat{\boldsymbol{\beta}}/\hat{\sigma}) \right] \hat{\beta}_j.$$

The scale factor is the average of $\hat{P}(y > 0|\mathbf{x})$ across the sample. (Unfortunately, the “margins” command in Stata reports the $\hat{\beta}_j$ and not the APEs.)

- Bootstrapping is convenient and feasible.
- Can compare the Tobit APEs on $E(y|\mathbf{x})$ to OLS estimates using entire sample. Also, can average across a subset of the covariates, fixing others at particular values.

• For discrete explanatory variables or for large changes in continuous ones, we can compute the difference in $E(y|\mathbf{x})$ at different values of $\mathbf{x}$. Suppose x_K is a binary variable (such as a policy indicator), and define, for each observation i, the two indices $\hat{w}_{i1} = \mathbf{x}_{i(K)}\hat{\boldsymbol{\beta}}_{(K)} + \hat{\beta}_K$ and $\hat{w}_{i0} = \mathbf{x}_{i(K)}\hat{\boldsymbol{\beta}}_{(K)}$, where $\mathbf{x}_{i(K)}$ is the $1 \times (K-1)$ row vector with x_{iK} dropped.Then, the average difference

$$N^{-1}\sum_{i=1}^{N}\{[\Phi(\hat{w}_{i1}/\hat{\sigma})\hat{w}_{i1} + \hat{\sigma}\phi(\hat{w}_{i1}/\hat{\sigma})] - [\Phi(\hat{w}_{i0}/\hat{\sigma})\hat{w}_{i0} + \hat{\sigma}\phi(\hat{w}_{i0}/\hat{\sigma})]\}$$

is the so-called *average treatment effect*.

• Different ways to define goodness-of-fit statistics. If we focus on $E(y|\mathbf{x})$, a simple one is the squared correlation between y_i and $\hat{E}(y_i|\mathbf{x}_i) = \Phi(\mathbf{x}_i\hat{\boldsymbol{\beta}}/\hat{\sigma})\mathbf{x}_i\hat{\boldsymbol{\beta}} + \hat{\sigma}\phi(\mathbf{x}_i\hat{\boldsymbol{\beta}}/\hat{\sigma})$. Or, can use a sum of squared residuals-type R-squared. Both are comparable to OLS R-squared. Or, we can use $\hat{E}(y_i|\mathbf{x}_i, y_i > 0) = \mathbf{x}_i\hat{\boldsymbol{\beta}} + \hat{\sigma}\lambda(\mathbf{x}_i\hat{\boldsymbol{\beta}}/\hat{\sigma})$ to look at the fit for nonlimit observations.

• If comparing different models with full distributions specified, can use log likelihood for the full sample. This measures the fit of the entire distribution. As always, we prefer the model with the highest log-likelihood function.

```
. sum hours

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       hours |       753    740.5764    871.3142          0       4950


. count if hours == 0
  325
```

```
. tobit hours nwifeinc educ exper expersq age kidslt6 kidsge6, ll(0)

Tobit regression                                  Number of obs   =        753
                                                  LR chi2(7)      =     271.59
                                                  Prob > chi2     =     0.0000
Log likelihood = -3819.0946                       Pseudo R2       =     0.0343

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -8.814243   4.459096    -1.98   0.048    -17.56811   -.0603724
        educ |   80.64561   21.58322     3.74   0.000     38.27453    123.0167
       exper |   131.5643   17.27938     7.61   0.000     97.64231    165.4863
     expersq |  -1.864158   .5376615    -3.47   0.001    -2.919667   -.8086479
         age |  -54.40501   7.418496    -7.33   0.000    -68.96862    -39.8414
     kidslt6 |  -894.0217   111.8779    -7.99   0.000    -1113.655   -674.3887
     kidsge6 |    -16.218   38.64136    -0.42   0.675    -92.07675    59.64075
       _cons |   965.3053   446.4358     2.16   0.031     88.88528    1841.725
-------------+----------------------------------------------------------------
      /sigma |   1122.022   41.57903                      1040.396    1203.647
------------------------------------------------------------------------------
  Obs. summary:        325  left-censored observations at hours<=0
                       428     uncensored observations
                         0 right-censored observations
```

```
. predict xbh, xb

. gen hoursh = normal(xbh/_b[/sigma])*xb + _b[/sigma]*normalden(xbh/_b[/sigma])

. sum hours hoursh

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       hours |       753    740.5764    871.3142          0       4950
      hoursh |       753    721.4201    473.6053   3.496456   1993.885

. corr hours hoursh
(obs=753)

             |    hours   hoursh
-------------+------------------
       hours |   1.0000
      hoursh |   0.5237   1.0000

. di .5237^2
.27426169

. * This squared correlation can be used as an R-squared to compare with other
. * models, including a linear model.
```

```
. reg hours hoursh, nocons

      Source |       SS       df       MS              Number of obs =     753
-------------+------------------------------           F(  1,   752) = 1031.68
       Model |   569084372     1   569084372           Prob > F      =  0.0000
    Residual |   414810722   752  551610.002           R-squared     =  0.5784
-------------+------------------------------           Adj R-squared =  0.5778
       Total |   983895094   753  1306633.59           Root MSE      =   742.7

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      hoursh |   1.007564    .031369    32.12   0.000     .9459831    1.069146
------------------------------------------------------------------------------

. * Note how close the coefficient is to one; this at least suggests the Tobit
. * provides sensible estimates of the "unconditional" mean.
```

```
. qui tobit hours nwifeinc educ exper expersq age kidslt6 kidsge6, ll(0)

. gen scale = normal(xbh/_b[/sigma])

. sum scale

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       scale |       753    .5886634    .2426614   .0092704    .960908

. di .589*80.65
47.50285

. di .589*(-894.02)
-526.57778
```

```
. do ex17_2_boot

. capture program drop tobit_boot

.
. program tobit_boot, rclass
  1.
. tobit hours nwifeinc educ exper expersq age kidslt6 kidsge6, ll(0)
  2.
. predict xbh, xb
  3. gen scale = normal(xbh/_b[/sigma])
  4. gen pe1=scale*_b[educ]
  5. sum pe1
  6. return scalar ape1=r(mean)
  7. gen xbh0 = xbh - _b[kidslt6]*kidslt6
  8. gen xbh1 = xbh0 + _b[kidslt6]
  9. gen xbh2 = xbh0 + _b[kidslt6]*2
 10. gen m2 = normal(xbh2/_b[/sigma])*xbh2 + _b[/sigma]*normalden(xbh2/_b[/sigma])
 11. gen m1 = normal(xbh1/_b[/sigma])*xbh1 + _b[/sigma]*normalden(xbh1/_b[/sigma])
 12. gen m0 = normal(xbh0/_b[/sigma])*xbh0 + _b[/sigma]*normalden(xbh1/_b[/sigma])
 13. gen pe2 = m1 - m0
 14. sum pe2
 15. return scalar ape2=r(mean)
 16. gen pe3 = m2 - m1
 17. sum pe3
 18. return scalar ape3=r(mean)
 19.
. drop xbh scale xbh1 xbh0 xbh2 m2 m1 m0 pe1 pe2 pe3
 20. end

.
. bootstrap r(ape1) r(ape2) r(ape3), reps(500) seed(123): tobit_boot
(running tobit_boot on estimation sample)

Bootstrap replications (500)
```

```
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200
..................................................   250
..................................................   300
..................................................   350
..................................................   400
..................................................   450
..................................................   500

Bootstrap results                               Number of obs      =       753
                                                Replications       =       500

      command:  tobit_boot
        _bs_1:  r(ape1)
        _bs_2:  r(ape2)
        _bs_3:  r(ape3)

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _bs_1 |   47.47311   13.04099     3.64   0.000     21.91324    73.03299
       _bs_2 |   -487.213   54.03773    -9.02   0.000     -593.125    -381.301
       _bs_3 |  -246.1717   12.31618   -19.99   0.000    -270.3109   -222.0324
------------------------------------------------------------------------------

.
. program drop tobit_boot

.
end of do-file
```

```
. tab kidslt6

 # kids < 6 |
      years |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        606       80.48       80.48
          1 |        118       15.67       96.15
          2 |         26        3.45       99.60
          3 |          3        0.40      100.00
------------+-----------------------------------
      Total |        753      100.00

. di 29 + 118
147

. di 118/147
.80272109

. di .8*487.2 + .2*246.2
439
```

```
. reg hours nwifeinc educ exper expersq age kidslt6 kidsge6, robust

Linear regression                                      Number of obs =     753
                                                       F(  7,   745) =   45.81
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2656
                                                       Root MSE      =  750.18

------------------------------------------------------------------------------
             |               Robust
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -3.446636   2.240662    -1.54   0.124    -7.845398    .9521268
        educ |   28.76112   13.03905     2.21   0.028     3.163468    54.35878
       exper |   65.67251   10.79419     6.08   0.000     44.48186    86.86316
     expersq |  -.7004939   .3720129    -1.88   0.060    -1.430812    .0298245
         age |  -30.51163   4.244791    -7.19   0.000    -38.84481   -22.17846
     kidslt6 |  -442.0899   57.46384    -7.69   0.000    -554.9002   -329.2796
     kidsge6 |  -32.77923   22.80238    -1.44   0.151     -77.5438    11.98535
       _cons |   1330.482   274.8776     4.84   0.000     790.8556    1870.109
------------------------------------------------------------------------------
```

```
. * Linear regression coefficient on kidslt6 is very close to the weighted
. * average of the Tobit effects.
. * But the linear regression coefficient for educ (28.8) is well below the APE
. * from the Tobit model (47.5).

. * The R-squared, about .266, is slightly below that for Tobit, about .274. The
. * difference is perhaps not as great as we expect. Remember, though, that
. * the OLS chooses the parameters to maximize the R-squared (minimize the sum of
. * squared residuals) while the Tobit estimates are chosen to maximize the log
. * likelihood. Nevertheless, the MLEs are, of course, good estimates, so the
. * difference in using MLE versus nonlinear least squares is probably minor.
```

4. SPECIFICATION ISSUES

Omitted Heterogeneity Independent of the Covariates

• The conclusions here are similar to the binary response case. If we add, say, q as unobserved heterogeneity, $y = \max(0, \mathbf{x}\boldsymbol{\beta} + \gamma q + u)$, $u|(\mathbf{x}, q) \sim Normal(0, \sigma^2)$, and q is also normally distributed and independent of $\mathbf{x}$, we can estimate $\boldsymbol{\beta}$ and $\gamma^2\tau^2 + \sigma^2$. Fortunately, the APEs only depend on $\gamma^2\tau^2 + \sigma^2$.

Heteroskedasticity

• Again, similar to probit. “Heteroskedastic Tobit” is a good way to extend functional form. Typically, $u|\mathbf{x} \sim Normal(0, \exp(2\mathbf{x}\boldsymbol{\delta})$ or restrict to a subset of $\mathbf{x}$. But it makes the partial effects on $E(y|\mathbf{x}, y > 0)$ and $E(y|\mathbf{x})$ more difficult to estimate.

• As in the probit case, there is a way to argue that the partial effects on the average structural function are the same sign as the β_j. See Wooldridge (2005, Festschrift for Rothenberg).

• Again, using “robust” with Tobit estimation does not produce consistent estimators of the $\boldsymbol{\beta}$ while somehow making the inference robust to heteroskedsticity in the underlying model.

Nonnormality

• Similar comments to heteroskedasticity. The usual Tobit MLE will not consistently estimate $\boldsymbol{\beta}$ (or σ), but it may yield reasonably close partial effects.

• Using a more flexible distribution for $D(u|\mathbf{x})$ might be a good idea, but one should not only compare estimated coefficients.

5. ESTIMATING PARAMETERS UNDER WEAKER ASSUMPTIONS

- When we assume y is generated as

$$y = \max(0, \mathbf{x}\boldsymbol{\beta} + u),$$

modeling $D(u|\mathbf{x})$ more flexibly than $Normal(0,\sigma^2)$ is a natural way to extend the model. At a minimum, it can be used to check the robustness of Tobit partial effects. This includes the two possibilities mentioned above: allow for heteroskedasticity in $Var(u|\mathbf{x})$ or nonnormality in $D(u|\mathbf{x})$.

• Another direction is to use an estimation method that identifies $\boldsymbol{\beta}$ without many restrictions on $D(u|\mathbf{x})$. We already saw that if $Med(u|\mathbf{x}) = Med(u) = 0$ then $Med(y|\mathbf{x}) = \max(0, \mathbf{x}\boldsymbol{\beta})$. This includes the Tobit model as a special case (even if we add heteroskedasticity, since $D(u|\mathbf{x})$ would still have median zero).

• Because LAD identifies the parameters of a correctly specified conditional median, we can solve

$$\min_{\mathbf{b} \in \mathbb{R}^K} \sum_{i=1}^{n} |y_i - \max(0, \mathbf{x}_i \mathbf{b})|$$

• This estimator is called the *censored least absolute deviations (CLAD)* estimator.

• The objective function is continuous in the parameters, so consistency is relatively straightforward provided we can establish identification (not always easy).

- The non-smoothness in the median function $\max(0, \mathbf{x}_i\mathbf{b})$ (as a function of $\mathbf{b}$) and of the LAD function make asymptotic normality hard, but Powell (1984) show it under general conditions. At least one x_j with nonzero β_j should be continuous, and the density of the error is continuous near zero and strictly positive at zero.
- Estimation of the asymptotic variance has been coded, too.

Sometimes bootstrapping is used.

• The estimates of the $\hat{\beta}_j$ give partial effects of continuous variables on the median once $\mathbf{x}\boldsymbol{\beta} > 0$. Ignoring what happens when $\mathbf{x}\boldsymbol{\beta} = 0$ [because he assumptions used to derive the asymptotic properties of LAD always include $P(\mathbf{x}\boldsymbol{\beta} = 0)$], we have

$$\frac{\partial Med(y|\mathbf{x})}{\partial x_j} = \beta_j 1[\mathbf{x}\boldsymbol{\beta} > 0].$$

• Given $\hat{\boldsymbol{\beta}}$ we can compute $\hat{\beta}_j 1[\mathbf{x}\hat{\boldsymbol{\beta}} > 0]$ for any vector $\mathbf{x}$.

• For a discrete change, must evaluate $\max(0, \mathbf{x}\hat{\boldsymbol{\beta}})$ at the different values of $\mathbf{x}$.

- Honoré (2008) has recently shown that even when u and $\mathbf{x}$ are not independent, the quantity

$$\beta_j P(y > 0)$$

can be interpreted as an average partial effect when x_j is continuous. Given the CLAD estimate $\hat{\beta}_j$, as before we can easily estimate this quantity by $\hat{\beta}_j \hat{\eta}$ where $\hat{\eta}$ is the fraction of strictly positive y_i.

- Unfortunately, this does not work for discrete changes in x_j.
- Plus we cannot use it to evaluate partial effects across different values of $\mathbf{x}$.

• One way to view the CLAD approach is that it identfies $Med(y|\mathbf{x}) = \max(0, \mathbf{x}\boldsymbol{\beta})$ for a variety of shapes of $D(u|\mathbf{x})$. But there is a cost: other features of $D(y|\mathbf{x})$, such as the mean, are not identified. So, CLAD does not allow us to aggregate the effects of a policy or program. We can get the median effect for groups indexed by the observed covariates.

• The CLAD approach does not work for two-part models, as we will see.

```
. clad hours nwifeinc educ exper expersq age kidslt6 kidsge6, ll(0) reps(500)

Initial sample size = 753
Final sample size = 504
Pseudo R2 = .12643936

Bootstrap statistics

Variable |   Reps   Observed       Bias   Std. Err.   [95% Conf. Interval]
---------+-------------------------------------------------------------------
nwifeinc |    500  -6.083408   -1.02711   6.052369   -17.97468  5.807859  (N)
---------+-------------------------------------------------------------------
    educ |    500   72.48654  -3.507613   32.45573    8.719817  136.2533  (N)
---------+-------------------------------------------------------------------
   exper |    500   120.1826   3.870493   21.22486    78.48153  161.8837  (N)
---------+-------------------------------------------------------------------
 expersq |    500  -1.403609  -.1508735   .7030393    -2.78489 -.0223267  (N)
---------+-------------------------------------------------------------------
     age |    500   -59.7946   -.102942   9.315551   -78.09713 -41.49206  (N)
---------+-------------------------------------------------------------------
 kidslt6 |    500  -1075.789  -14.31295   215.2858   -1498.767 -652.8106  (N)
---------+-------------------------------------------------------------------
 kidsge6 |    500  -97.81434  -20.24035   47.63005   -191.3945 -4.234172  (N)
---------+-------------------------------------------------------------------
   const |    500   1568.482   58.27831   589.2338    410.7975  2726.167  (N)
-----------------------------------------------------------------------------
```

6. ENDOGENOUS EXPLANATORY VARIABLES

- Can use a linear model and view the 2SLS estimates (say) as (rough?) estimates of average partial effects. Has the benefit of "working" for any kind of y_2.
- If we want to allow nonconstant partial effects, we need to turn to nonlinear models.
- With a single EEV (for simplicity), consider the model

$$\begin{aligned} y_1 &= \max(0, \alpha_1 y_2 + \mathbf{z}_1 \boldsymbol{\delta}_1 + u_1) \\ u_1 | \mathbf{z} &\sim Normal(0, \sigma_1^2) \end{aligned}$$

where $\mathbf{z}$ is the vector of all endogenous variables.

- Analysis goes through if we replace $(\mathbf{z}_1, y_2)$ with any known function $\mathbf{x}_1 \equiv \mathbf{g}_1(\mathbf{z}_1, y_2)$. Might be useful to include squares and cross products, y_2^2 and $y_2\mathbf{z}_1$, to make the functional form more flexible.
- The parameters $(\alpha_1, \boldsymbol{\delta}_1, \sigma_1^2)$ index the average structural function, and so they index the APEs, too.
- The Smith-Blundell (1986) approach is a control function approach.

$$y_2 = \mathbf{z}\boldsymbol{\delta}_2 + v_2 = \mathbf{z}_1\boldsymbol{\delta}_{21} + \mathbf{z}_2\boldsymbol{\delta}_{22} + v_2,\ \boldsymbol{\delta}_{22} \neq \mathbf{0}$$

$$v_2|\mathbf{z} \sim Normal(0, \tau_2^2)$$

- Can relax normality in two-step methods. In fact, sufficient is

$$u_1 = \theta_1 v_2 + e_1$$

where $\theta_1 = Cov(v_2, u_1)/\tau_2^2$ and e_1 given $(v_2, \mathbf{z})$ has a $Normal(0, \tau_1^2)$ distribution with $\tau_1^2 = \sigma_1^2 - \theta_1^2 \tau_2^2$.

- As before, the CF approach is a two-step method. Write

$$y_1 = \max(0, \alpha_1 y_2 + \mathbf{z}_1 \boldsymbol{\delta}_1 + \theta_1 v_2 + e_1)$$

which follows a Tobit model with explanatory variables $(y_2, \mathbf{z}_1, v_2)$. It is useful to write this as

$$\begin{aligned} D(y_1|y_2, \mathbf{z}) &= Tobit(\alpha_1 y_2 + \mathbf{z}_1 \boldsymbol{\delta}_1 + \theta_1 (y_2 - \mathbf{z}\boldsymbol{\delta}_2), \tau_1^2) \\ &= Tobit(\alpha_1 y_2 + \mathbf{z}_1 \boldsymbol{\delta}_1 + \theta_1 v_2, \tau_1^2) \end{aligned}$$

• So (1) Run the OLS regression y_{i2} on $\mathbf{z}_i$ and get the residuals, $\hat{v}_{i2}$. (2) Run Tobit of y_{i1} on $(y_{i2}, \mathbf{z}_{i1}, v_{i2})$.

• The second step gives $\hat{\alpha}_1, \hat{\boldsymbol{\delta}}_1, \hat{\theta}_1$, and $\hat{\tau}_1^2$. If we want to use the original parameters to compute APEs, set $\hat{\sigma}_1^2 = \hat{\theta}_1 \hat{\tau}_2^2 + \hat{\tau}_1^2$ where $\hat{\tau}_2^2$ is the estimated reduced form variance.

- As a useful shorthand, define the Tobit mean function as

$$m(a,\sigma^2) \equiv \Phi(a/\sigma)a + \sigma\phi(a/\sigma).$$

- Then the estimated partial effects can be obtained by computing derivatives or differences of $m(\mathbf{z}_1\hat{\boldsymbol{\delta}}_1 + \hat{\alpha}_1 y_2, \hat{\sigma}_1^2)$ with respect to elements of $(\mathbf{z}_1, y_2)$, just as we did in the case of exogenous explanatory variables.

- Or, we can use

$$\widehat{ASF}(y_2,\mathbf{z}_1) = N^{-1}\sum_{i=1}^{N} m(\hat{\alpha}_1 y_2 + \mathbf{z}_1\hat{\boldsymbol{\delta}}_1 + \hat{\theta}_1\hat{v}_{i2}, \hat{\tau}_1^2)$$

which is valid under weaker assumptions.

• For example, the partial effect with respect to y_2 is estimated as

$$\left\{N^{-1}\sum_{i=1}^{N}\Phi[(\hat{\alpha}_1 y_2 + \mathbf{z}_1\hat{\boldsymbol{\delta}}_1 + \hat{\theta}_1\hat{v}_{i2})/\hat{\tau}_1]\right\}\hat{\alpha}_1,$$

which depends on $(y_2, \mathbf{z}_1)$. Can do a further averaging out. to get a single number.

• As with all control function procedures, we can easily allow more general functional forms in both the exogenous and endogenous variables (such as squares and interactions). In fact, if we replace $\mathbf{x}_1 = (\mathbf{z}_1, y_2)$ with $\mathbf{x}_1 = \mathbf{g}_1(\mathbf{z}_1, y_2)$, then the estimation procedure and calculation of the ASF are unchanged.

• Some additional flexibility is gained by allowing $E(u_1|v_2)$ to be nonlinear – for example, a quadratic function, $E(u_1|v_2) = \theta_1 v_2 + \psi_1(v_2^2 - \tau_2^2)$ (where the variance of v_2 is subtracted from v_2^2 to ensure $E(u_1) = 0$). Then we can write

$$u_1 = \theta_1 v_2 + \psi_1(v_2^2 - \tau_2^2) + e_1$$

and, if we maintain that e_1 is independent of v_2 and normally distributed, we can easily extend the two-step procedure: the second step of the procedure adds $\hat{v}_2$ and $\hat{v}_2^2 - \hat{\tau}_2^2$ as the control functions in the Tobit estimation.

• The easiest way to obtain APEs in this case is to use derivatives and changes with respect to elements of $\mathbf{x}_1$ of

$$N^{-1}\sum_{i=1}^{N} m(\mathbf{x}_1\hat{\boldsymbol{\beta}}_1 + \hat{\theta}_1\hat{v}_{i2} + \hat{\psi}_1(\hat{v}_{i2}^2 - \hat{\tau}_2^2), \hat{\tau}_1^2),$$

• In the case where (u_1, v_2) is independent of $\mathbf{z}$ with a bivariate normal distribution, we can use MLE. Again, we use $f(y_1, y_2|\mathbf{z}) = f(y_1|y_2, \mathbf{z})f(y_2|\mathbf{z})$, and note that

$$\begin{aligned} D(y_1|y_2, \mathbf{z}) &= Tobit(\alpha_1 y_2 + \mathbf{z}_1\boldsymbol{\delta}_1 + (\eta_1/\tau_2^2)(y_2 - \mathbf{z}\boldsymbol{\delta}_2), \sigma_1^2 - (\eta_1^2/\tau_2^2)) \\ D(y_2|\mathbf{z}) &= Normal(\mathbf{z}\boldsymbol{\delta}_2, \tau_2^2) \end{aligned}$$

where $\eta_1 = Cov(v_2, u_1)$.

- In Stata, the command for joint MLE is “ivtobit.”
- Even if you use MLE, might want bootstrap standard errors for partial effects; otherwise the delta method is complicated.
- Clearly need an exclusion restriction for identification
, just as in the linear case.

```
. * Labor supply with nwifeinc potentially endogenous.
. * First estimate reduced form, where husband's education
. * is the IV for nwifeinc:

. reg nwifeinc huseduc educ exper expersq age kidslt6 kidsge6

      Source |       SS       df       MS              Number of obs =     753
-------------+------------------------------           F(  7,   745) =   27.13
       Model |  20676.7705     7  2953.82436           Prob > F      =  0.0000
    Residual |  81120.3451   745  108.886369           R-squared     =  0.2031
-------------+------------------------------           Adj R-squared =  0.1956
       Total |  101797.116   752  135.368505           Root MSE      =  10.435

------------------------------------------------------------------------------
    nwifeinc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     huseduc |   1.178155   .1609449     7.32   0.000     .8621956    1.494115
        educ |   .6746951   .2136829     3.16   0.002     .2552029    1.094187
       exper |  -.3129877   .1382549    -2.26   0.024    -.5844034   -.0415721
     expersq |  -.0004776   .0045196    -0.11   0.916    -.0093501     .008395
         age |   .3401521   .0597084     5.70   0.000     .2229354    .4573687
     kidslt6 |   .8262719   .8183785     1.01   0.313    -.7803305    2.432874
     kidsge6 |   .4355289   .3219888     1.35   0.177    -.1965845    1.067642
       _cons |  -14.72048   3.787326    -3.89   0.000    -22.15559   -7.285383
------------------------------------------------------------------------------
```

```
. predict v2h, resid

. tobit hours nwifeinc v2h educ exper expersq age kidslt6 kidsge6, ll(0)

Tobit regression                                  Number of obs   =        753
                                                  LR chi2(8)      =     273.76
                                                  Prob > chi2     =     0.0000
Log likelihood = -3818.0118                       Pseudo R2       =     0.0346

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -31.48215    16.0376    -1.96   0.050    -62.96641    .0021189
         v2h |   24.41832   16.58452     1.47   0.141    -8.139637    56.97628
        educ |   116.7814   32.75978     3.56   0.000     52.46891    181.0939
       exper |   124.3488   17.87502     6.96   0.000     89.25736    159.4402
     expersq |    -1.8972   .5371614    -3.53   0.000     -2.95173   -.8426702
         age |  -46.89244   8.957672    -5.23   0.000    -64.47773   -29.30716
     kidslt6 |  -867.9131   112.9024    -7.69   0.000    -1089.558   -646.2684
     kidsge6 |   -6.32605   39.16561    -0.16   0.872    -83.21414    70.56204
       _cons |   722.1032    475.689     1.52   0.129    -211.7472    1655.954
-------------+----------------------------------------------------------------
      /sigma |   1119.844   41.49319                      1038.387    1201.302
------------------------------------------------------------------------------
  Obs. summary:        325  left-censored observations at hours<=0
                       428     uncensored observations
                         0 right-censored observations
```

```
. * Some, but not strong, evidence of endogeneity of nwifeinc: p-value = .141.

. * Now use MLE. Should get same estimates because
. * just identified.

. ivtobit hours educ exper expersq age kidslt6 kidsge6 (nwifeinc = huseduc),
          ll(0)

Tobit model with endogenous regressors             Number of obs   =        753
                                                   Wald chi2(7)    =     248.26
Log likelihood = -6648.3509                        Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hours        |
    nwifeinc |  -31.48201   16.37719    -1.92   0.055    -63.58071    .6166817
        educ |   116.7812   33.45288     3.49   0.000     51.21473    182.3476
       exper |   124.3488   18.21246     6.83   0.000     88.65303    160.0446
     expersq |    -1.8972   .5482816    -3.46   0.001    -2.971812   -.8225878
         age |  -46.89249   9.135658    -5.13   0.000    -64.79805   -28.98692
     kidslt6 |  -867.9132   114.6915    -7.57   0.000    -1092.705   -643.1219
     kidsge6 |  -6.326111   39.96683    -0.16   0.874    -84.65966    72.00744
       _cons |   722.1047   485.6203     1.49   0.137    -229.6936    1673.903
-------------+----------------------------------------------------------------
      /alpha |   24.41818   16.91313     1.44   0.149    -8.730942     57.5673
        /lns |   7.020945   .0370527   189.49   0.000     6.948323    7.093567
        /lnv |   2.339812   .0257684    90.80   0.000     2.289307    2.390317
-------------+----------------------------------------------------------------
           s |   1119.844   41.49329                      1041.401    1204.195
           v |   10.37929   .2674576                      9.868095    10.91696
------------------------------------------------------------------------------
Instrumented:  nwifeinc
Instruments:   educ exper expersq age kidslt6 kidsge6 huseduc
------------------------------------------------------------------------------
```

```
Wald test of exogeneity (/alpha = 0): chi2(1) =     2.08  Prob > chi2 = 0.1488

  Obs. summary:        325  left-censored observations at hours<=0
                       428     uncensored observations

. * Clearly some rounding error, but the same.

. * Should compute APEs, compare with linear IV.
```

7. PANEL DATA

- Guess what? It is not crazy to use a linear unobserved effects model when y_{it} is the response variable of interest.
- Drawbacks are the functional form for $E(y_{it}|\mathbf{x}_{it}, c_i)$, but we need not restrict $D(c_i|\mathbf{x}_i)$ – just as in any limited dependent variable context.
- Start with the unobserved effects Tobit model

$$\begin{aligned} y_{it} &= \max(0, \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}) \\ D(u_{it}|\mathbf{x}_{it}, c_i) &= Normal(0, \sigma_u^2) \end{aligned}$$

- Ideally, we could proceed without further assumptions, although that would leave open the question of values to insert for c in, say, partial effects on $E(y_{it}|\mathbf{x}_{it} = \mathbf{x}_t, c_i = c)$, or how to average out across the distribution of c_i.
- If the c_i are treated as parameters to estimate along with $\boldsymbol{\beta}$ and σ_u^2, for "small" T an incidental parameters problem arises with for estimating $\boldsymbol{\beta}$ and σ_u^2. It is possible the average partial effects are better behaved than the estimates themselves, but I have no evidence on that.

• What parametric assumptions are sufficient for estimating $\boldsymbol{\beta}$, σ_u^2, and the APEs?

• Not surprisingly, the same assumptions that come up in UE probit models arise here as well.

• Strict exogeneity conditional on c_i:

$$D(u_{it}|\mathbf{x}_i, c_i) = D(u_{it}|\mathbf{x}_{it}, c_i)$$

• Conditional independence (conditional on $\mathbf{x}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})$ and c_i):

$$D(u_{i1}, \ldots, u_{iT}|\mathbf{x}_i, c_i) = D(u_{i1}|\mathbf{x}_i, c_i) \cdots D(u_{iT}|\mathbf{x}_i, c_i)$$

• Model for $D(c_i|\mathbf{x}_i)$ (Mundlak special case of Chamberlain approach):

$$c_i = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i,\ a_i|\mathbf{x}_i \sim Normal(0, \sigma_a^2).$$

• Can include time dummies in $\mathbf{x}_{it}$ but omit from $\bar{\mathbf{x}}_i$. Can also include time-constant elements (say $\mathbf{w}_i$) as controls. That is, write

$c_i = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + \mathbf{w}_i\boldsymbol{\zeta} + a_i$.

• If $\boldsymbol{\xi} = \mathbf{0}$, get the traditional random effects Tobit model. Adding $\bar{\mathbf{x}}_i\boldsymbol{\xi}$ allows a specific form of correlation.

• MLE (conditional on $\mathbf{x}_i$) is relatively straightforward. It is based on the joint distribution $D(y_{i1}, \ldots, y_{iT}|\mathbf{x}_i)$.

- As with CRE probit, get log likelihood by “integrating out” c_i:

$$l_i(\boldsymbol{\beta}, \psi, \boldsymbol{\xi}, \sigma_a^2, \sigma_u^2) = \log\left[\int_{-\infty}^{\infty}\left(\prod_{t=1}^{T} f(y_{it}|\mathbf{x}_{it}, c; \boldsymbol{\beta}, \sigma_u^2)\right) h(c|\bar{\mathbf{x}}_i; \psi, \boldsymbol{\xi}, \sigma_a^2) dc\right]$$

where $f(y_t|\mathbf{x}_t, c; \boldsymbol{\beta}, \sigma_u^2)$ is the $Tobit(\mathbf{x}_t\boldsymbol{\beta} + c, \sigma_u^2)$ density and $h(c|\bar{\mathbf{x}}; \psi, \boldsymbol{\xi}, \sigma_a^2)$ is the Normal density.

- In Stata, called "xttobit" with the "re" option:

```
xttobit y x1 x2 ... xK x1bar ... xKbar, ll(0) re
```

- As in the probit case, we can estimate μ_c and σ_c^2:

$$\hat{\mu}_c = \hat{\psi} + \bar{\mathbf{x}}\hat{\boldsymbol{\xi}}$$

$$\hat{\sigma}_c^2 \equiv \hat{\boldsymbol{\xi}}' \left(N^{-1} \sum_{i=1}^{N} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})'(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}) \right) \hat{\boldsymbol{\xi}} + \hat{\sigma}_a^2$$

• Then, we can evaluate the partial effects of the Tobit function, $m(\mathbf{x}_t\hat{\boldsymbol{\beta}} + c, \hat{\sigma}_u^2)$ at different values of c, including $\hat{\mu}_c$ and $\hat{\mu}_c \pm k\hat{\sigma}_c$.

• Take derivatives or changes with respect to $\mathbf{x}_t$. For a continuous variable,

$$\hat{\beta}_j \Phi[(\mathbf{x}_t\hat{\boldsymbol{\beta}} + c)/\hat{\sigma}_u]$$

• APEs can be estimated from the mean function for the Tobit:

$$\widehat{ASF}(\mathbf{x}_t) = N^{-1}\sum_{i=1}^{n} m(\mathbf{x}_t\hat{\boldsymbol{\beta}} + \hat{\psi} + \bar{\mathbf{x}}_i\hat{\boldsymbol{\xi}}, \hat{\sigma}_a^2 + \hat{\sigma}_u^2)$$

• Take derivatives and differences with respect to elements of $\mathbf{x}_t$; can further average. Panel bootstrap!

- For a continuous x_{tj},

$$\hat{\beta}_j \left[N^{-1} \sum_{i=1}^{N} \Phi[(\mathbf{x}_t \hat{\boldsymbol{\beta}} + \hat{\psi} + \bar{\mathbf{x}}_i \hat{\boldsymbol{\xi}})/(\hat{\sigma}_a^2 + \hat{\sigma}_u^2)^{1/2}] \right]$$

- To estimate the APEs, note it suffices to estimate $\sigma_v^2 = \sigma_a^2 + \sigma_u^2$.
- Suppose we drop the conditional independence assumption and allow

and serial dependence in $\{u_{it}\}$. then

$$D(y_{it}|\mathbf{x}_i) = D(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i) = Tobit(\mathbf{x}_{it}\boldsymbol{\beta} + \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi}, \sigma_v^2)$$

• So, we can apply pooled Tobit, ignoring the serial correlation, to estimate $\boldsymbol{\beta}, \psi, \boldsymbol{\xi}$, and σ_v^2. We can then use the above formula for the APEs. We cannot estimate PEAs because $E(c_i)$ is not identified; neither is $\boldsymbol{\beta}$ nor σ_u^2.

• Adding time constant variables $\mathbf{w}_i$ to the full MLE or the pooled Tobit is straightforward. Interpreting the coefficients is more difficult. Are they in the model because they are correalted with c_i, or do they appear directly? One cannot make this distinction.

```
. use psid80_92

. tab year

   80 to 92 |      Freq.     Percent        Cum.
------------+-----------------------------------
         80 |        898        7.69        7.69
         81 |        898        7.69       15.38
         82 |        898        7.69       23.08
         83 |        898        7.69       30.77
         84 |        898        7.69       38.46
         85 |        898        7.69       46.15
         86 |        898        7.69       53.85
         87 |        898        7.69       61.54
         88 |        898        7.69       69.23
         89 |        898        7.69       76.92
         90 |        898        7.69       84.62
         91 |        898        7.69       92.31
         92 |        898        7.69      100.00
------------+-----------------------------------
      Total |     11,674      100.00

. * First, linear FE:

. xtreg hours nwifeinc ch0_2 ch3_5 ch6_17 marr y81-y92, fe cluster(id)

Fixed-effects (within) regression               Number of obs      =     11674
Group variable (i): id                          Number of groups   =       898

R-sq:  within  = 0.0719                         Obs per group: min =        13
       between = 0.0936                                        avg =      13.0
       overall = 0.0855                                        max =        13

                                                F(17,11657)        =     15.72
corr(u_i, Xb)  = -0.0945                        Prob > F           =    0.0000
```

```
                                   (Std. Err. adjusted for 898 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.7752375   .3429502    -2.26   0.024    -1.448316   -.1021593
       ch0_2 |  -342.3774   26.64763   -12.85   0.000    -394.6763   -290.0784
       ch3_5 |  -254.1283   25.87788    -9.82   0.000    -304.9165     -203.34
      ch6_17 |  -42.95787   14.88673    -2.89   0.004    -72.17475   -13.74099
        marr |  -634.8048   286.1714    -2.22   0.027    -1196.448    -73.1613
         y81 |  -4.819715   16.29731    -0.30   0.767    -36.80502    27.16559
         y82 |  -14.88765    21.1851    -0.70   0.482     -56.4658    26.69049
         y83 |   6.612531   22.49192     0.29   0.769    -37.53039    50.75545
         y84 |   93.79139   25.58646     3.67   0.000      43.5751    144.0077
         y85 |   88.73714   25.97019     3.42   0.001     37.76773    139.7065
         y86 |   82.66214   27.36886     3.02   0.003     28.94769    136.3766
         y87 |   64.28464   27.83649     2.31   0.021     9.652411    118.9169
         y88 |   63.79163   29.35211     2.17   0.030     6.184826    121.3984
         y89 |   72.98518   30.60838     2.38   0.017     12.91279    133.0576
         y90 |   71.24956   31.55331     2.26   0.024     9.322657    133.1765
         y91 |   64.67996   32.47097     1.99   0.047     .9520418    128.4079
         y92 |   16.01242   33.21255     0.48   0.630    -49.17093    81.19577
       _cons |    1786.02    247.297     7.22   0.000     1300.672    2271.368
-------------+----------------------------------------------------------------
     sigma_u |  701.66249
     sigma_e |  503.92334
         rho |  .65972225   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. * Compute time averages:

. egen nwifeincb = mean(nwifeinc), by(id)

. egen ch0_2b = mean(ch0_2), by(id)
```

```
. egen ch3_5b = mean(ch3_5), by(id)

. egen ch6_17b = mean(ch6_17), by(id)

. egen marrb = mean(marr), by(id)

. * Correlated RE Tobit:

. xttobit hours nwifeinc ch0_2 ch3_5 ch6_17 marr y81-y92 nwifeincb-marrb, ll(0)

Random-effects tobit regression                 Number of obs      =     11674
Group variable (i): id                          Number of groups   =       898

Random effects u_i ~Gaussian                   Obs per group: min =        13
                                                              avg =      13.0
                                                              max =        13

                                                Wald chi2(22)      =   1501.20
Log likelihood  = -70733.195                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -1.554228   .3816927    -4.07   0.000    -2.302332   -.8061243
       ch0_2 |   -472.088   23.03087   -20.50   0.000    -517.2277   -426.9483
       ch3_5 |  -329.3896   19.49411   -16.90   0.000    -367.5974   -291.1819
      ch6_17 |  -46.11619   10.89609    -4.23   0.000    -67.47213   -24.76024
        marr |  -784.1809   155.0133    -5.06   0.000    -1088.001   -480.3604
         y81 |  -7.060588   31.52257    -0.22   0.823    -68.84369    54.72251
         y82 |   -38.9034   31.70009    -1.23   0.220    -101.0344    23.22764
         y83 |  -9.719573   31.68694    -0.31   0.759    -71.82483    52.38569
         y84 |   99.77618   31.61932     3.16   0.002     37.80345    161.7489
         y85 |   89.15912    31.7439     2.81   0.005     26.94222     151.376
         y86 |   82.60212   31.76385     2.60   0.009     20.34612    144.8581
```

```
         y87 |   48.59097   31.98439     1.52   0.129    -14.09729    111.2792
         y88 |   53.52189   32.09804     1.67   0.095    -9.389108    116.4329
         y89 |   68.69013   32.23667     2.13   0.033     5.507414    131.8728
         y90 |    71.2654    32.3657     2.20   0.028       7.8298     134.701
         y91 |   64.89096   32.48217     2.00   0.046     1.227067    128.5548
         y92 |   4.334129   32.82961     0.13   0.895    -60.01072    68.67898
   nwifeincb |  -7.639696   .6815067   -11.21   0.000    -8.975424   -6.303967
      ch0_2b |  -143.4709   155.0915    -0.93   0.355    -447.4448    160.5029
      ch3_5b |   531.2027    150.388     3.53   0.000     236.4475    825.9578
     ch6_17b |   5.854889   28.04159     0.21   0.835    -49.10563     60.8154
       marrb |   422.1631    161.491     2.61   0.009     105.6465    738.6796
       _cons |   1646.362   45.26091    36.37   0.000     1557.652    1735.072
-------------+----------------------------------------------------------------
    /sigma_u |   756.4032   10.45016    72.38   0.000     735.9213    776.8851
    /sigma_e |   621.7044    5.02536   123.71   0.000     611.8549    631.5539
-------------+----------------------------------------------------------------
         rho |   .5968169   .0069011                      .5832357    .6102823
------------------------------------------------------------------------------

  Observation summary:      3071  left-censored observations
                            8603     uncensored observations
                               0 right-censored observations

. testparm nwifeincb-marrb

 ( 1)  [hours]nwifeincb = 0
 ( 2)  [hours]ch0_2b = 0
 ( 3)  [hours]ch3_5b = 0
 ( 4)  [hours]ch6_17b = 0
 ( 5)  [hours]marrb = 0

           chi2(  5) =  165.08
         Prob > chi2 =    0.0000
```

```
. gen xbhata = xbhat/sqrt(756.4032^2 +  621.7044^2)

. gen PHIhata = norm(xbhata)

. sum PHIhata if y92

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     PHIhata |       898    .8367103    .0953704   .0029178   .9654008

------------------------------------------------------------------------------


. di  (.837)*(-1.554)
-1.300698

. di  (.837)*(-472.09)
-395.13933

. * Pooled Tobit with Time Averages:

. tobit hours nwifeinc ch0_2 ch3_5 ch6_17 marr y81-y92 nwifeincb-marrb, ll(0)

Tobit regression                                  Number of obs   =      11674
                                                  LR chi2(22)     =    1352.20
                                                  Prob > chi2     =     0.0000
Log likelihood = -75313.315                       Pseudo R2       =     0.0089

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -1.796524   .6073205    -2.96   0.003    -2.986975   -.6060744
       ch0_2 |  -491.6069   38.36112   -12.82   0.000    -566.8011   -416.4127
       ch3_5 |  -347.5099    32.7817   -10.60   0.000    -411.7675   -283.2523
      ch6_17 |  -48.12398   18.14746    -2.65   0.008    -83.69604   -12.55191
```

```
        marr |  -788.6605   257.0461    -3.07   0.002    -1292.514   -284.8071
         y81 |  -1.723103   52.74963    -0.03   0.974    -105.1212     101.675
         y82 |  -29.93459   52.90393    -0.57   0.572    -133.6352    73.76597
         y83 |   .1544965   52.88423     0.00   0.998    -103.5075    103.8165
         y84 |   111.7593   52.84133     2.11   0.034     8.181439    215.3372
         y85 |    98.8203   53.02693     1.86   0.062    -5.121366     202.762
         y86 |   91.11779   53.07409     1.72   0.086    -12.91632    195.1519
         y87 |   56.20641   53.35906     1.05   0.292    -48.38629    160.7991
         y88 |   58.45143   53.59859     1.09   0.275    -46.61078    163.5136
         y89 |   74.11085   53.83913     1.38   0.169    -31.42287    179.6446
         y90 |   77.83721   54.05111     1.44   0.150    -28.11203    183.7865
         y91 |   70.43439   54.27841     1.30   0.194    -35.96039    176.8292
         y92 |   4.969863   54.81622     0.09   0.928    -102.4791    112.4188
    nwifeincb |  -7.248981   .7293248    -9.94   0.000    -8.678579   -5.819382
      ch0_2b |   152.0109   124.2391     1.22   0.221    -91.51857    395.5403
      ch3_5b |   151.7502   118.9341     1.28   0.202    -81.38056     384.881
     ch6_17b |   44.11858   25.07548     1.76   0.079    -5.033552    93.27072
       marrb |   471.4367   259.4683     1.82   0.069    -37.16466    980.0381
       _cons |   1581.923   46.08447    34.33   0.000      1491.59    1672.256
-------------+----------------------------------------------------------------
      /sigma |   1079.331   8.836301                       1062.01    1096.651
------------------------------------------------------------------------------
  Obs. summary:       3071  left-censored observations at hours<=0
                      8603     uncensored observations
                         0 right-censored observations


. * These differ somewhat, but not in major ways, from the full MLEs.

. * Now drop the time averages, so RE Tobit:

. xttobit hours nwifeinc ch0_2 ch3_5 ch6_17 marr y81-y92, ll(0)

Random-effects tobit regression                 Number of obs      =     11674
Group variable (i): id                          Number of groups   =       898
```

Random effects u_i ~Gaussian

Obs per group:	min	=	13
	avg	=	13.0
	max	=	13

Wald chi2(17) = 1222.37

Log likelihood = -70782.086

Prob > chi2 = 0.0000

hours	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
nwifeinc	-2.25119	.3248083	-6.93	0.000	-2.887803	-1.614578
ch0_2	-459.927	22.67389	-20.28	0.000	-504.3671	-415.487
ch3_5	-313.4996	18.81897	-16.66	0.000	-350.3841	-276.6151
ch6_17	-32.33052	9.819359	-3.29	0.001	-51.57611	-13.08493
marr	-657.5755	48.93306	-13.44	0.000	-753.4825	-561.6684
y81	-6.015057	31.64666	-0.19	0.849	-68.04136	56.01125
y82	-37.89952	31.82432	-1.19	0.234	-100.274	24.47499
y83	-7.2714	31.78778	-0.23	0.819	-69.5743	55.0315
y84	104.3436	31.71544	3.29	0.001	42.18249	166.5047
y85	94.90622	31.82266	2.98	0.003	32.53496	157.2775
y86	89.38999	31.84555	2.81	0.005	26.97386	151.8061
y87	57.1533	32.03317	1.78	0.074	-5.630564	119.9372
y88	64.08813	32.11484	2.00	0.046	1.144192	127.0321
y89	81.55682	32.20542	2.53	0.011	18.43536	144.6783
y90	85.75216	32.26838	2.66	0.008	22.50728	148.997
y91	80.93763	32.36379	2.50	0.012	17.50576	144.3695
y92	22.68549	32.63686	0.70	0.487	-41.28158	86.65255
_cons	1676.368	39.27514	42.68	0.000	1599.39	1753.346
/sigma_u	768.5483	12.40411	61.96	0.000	744.2367	792.8599
/sigma_e	624.285	5.068197	123.18	0.000	614.3515	634.2185
rho	.6024761	.0077085			.5872944	.6175041

```
  Observation summary:      3071  left-censored observations
                            8603     uncensored observations
                               0 right-censored observations

. predict xbhat, xb

. gen xbhata = xbhat/sqrt(768.5483^2 +  624.285^2)

. gen PHIhata = normal(xbhata)

. sum PHIhata if y92

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     PHIhata |       898    .8240658    .0724009   .3761031   .9578886

. * The scale factor is similar, but the coefficient estimates are somewhat different.
```

- There is also a way to estimate dynamic Tobit models using RE software.

- Assume, as with dynamic probit and ordered probit,

$$D(y_{it}|\mathbf{z}_i, y_{i,t-1}, \ldots, y_{i0}, c_i) = D(y_{it}|\mathbf{z}_{it}, y_{i,t-1}, c_i),\ t = 1, \ldots, T,$$

which combines correct dynamic specification with strict exogeneity of $\{\mathbf{z}_{it}\}$.

• Tobit model: We assume

$$y_{it} = \max(0, \mathbf{z}_{it}\boldsymbol{\delta} + \rho y_{i,t-1} + c_i + u_{it}), t = 1, \ldots, T$$

and

$$u_{it}|(\mathbf{z}_i, y_{i,t-1}, \ldots, y_{i0}, c_i) \sim Normal(0, \sigma_u^2), t = 1, \ldots, T.$$

• Makes sense only for corner solutions. Even then, not clear how lagged y should appear. Could define a dummy variable $w_{it} = 1[y_{it} = 1]$ and use, say, $\rho_1(1 - w_{i,t-1}) + \rho_2 y_{i,t-1}$. Can also interact these with the $\mathbf{z}_{it}$.

• A simple analysis is obtained from

$$c_i|\mathbf{z}_i, y_{i0} \sim Normal(\psi + \xi_0 y_{i0} + \mathbf{z}_i\boldsymbol{\xi}, \sigma_a^2)$$

or by letting this depend more flexibly on the initial value, y_{i0}, as with the lags above. Then we have

$$y_{it} = \max(0, \mathbf{z}_{it}\boldsymbol{\delta} + \rho y_{i,t-1} + \psi + \xi_0 y_{i0} + \mathbf{z}_i\boldsymbol{\xi} + a_i + u_{it}).$$

• The log-likelihood takes the same form as the RE Tobit model with strictly exogenous variables, even though the explanatory variables are not strictly exogenous.

- As in the probit and OP cases, a pooled Tobit analysis, with the same set of explanatory variables, is not consistent for the parameters.
- The APEs are again easy to compute:

$$N^{-1}\sum_{i=1}^{N} m(\mathbf{z}_t\hat{\boldsymbol{\delta}} + \hat{\rho}y_{t-1} + \hat{\psi} + \hat{\xi}_o y_{i0} + \mathbf{z}_i\hat{\boldsymbol{\xi}}, \hat{\sigma}_a^2 + \hat{\sigma}_u^2),$$

where all estimates are from the MLE procedure.

- For a continuous variable, the scale factor is

$$N^{-1}\sum_{i=1}^{N} \Phi[(\mathbf{z}_t\hat{\boldsymbol{\delta}} + \hat{\rho}y_{t-1} + \hat{\psi} + \hat{\xi}_o y_{i0} + \mathbf{z}_i\hat{\boldsymbol{\xi}})(\hat{\sigma}_a^2 + \hat{\sigma}_u^2)^{-1/2}],$$

and one can further average across $(\mathbf{z}_t, y_{t-1})$.

• Extensions along the lines of allowing heteroskedasticity in $D(c_i|y_{i0},\mathbf{z}_i)$, flexible conditional means, and even more flexible distributions, seem worth exploring.

Honoré (1993) shows how to estimate $\boldsymbol{\delta}$ and ρ without distributional assumptions for c_i or u_{it}. Partial effects at different values of $(\mathbf{z}_t, y_{t-1})$ are not available, and y_{t-1} must appear in linear, additive form. (It is easy to extend the CMLE approach allow general functions of $(\mathbf{z}_{it}, y_{i,t-1})$.)

8. TWO-LIMIT TOBIT MODELS

- Now allow for two limits. These might be logical or institutional constraints. Common are corners at 0 and 1 or 0 and 100.
- But suppose workers are allowed to contribute at most 15% of their earnings to a tax-deferred pension plan, and y_i is the percentage of income contributed for worker i, then the corners are at zero and 15. (The arbitrariness of the cap of 15 raises other interesting questions, such as: what would happen if the cap were not there, or what would happen if it is raised?)

• Generally, let $a_1 < a_2$ be the two limit values of y in the population. Then the *two-limit Tobit* model is most easily defined in terms of an underlying latent variable:

$$
\begin{aligned}
y^* &= \mathbf{x}\boldsymbol{\beta} + u, \;\; u|\mathbf{x} \sim \mathit{Normal}(0, \sigma^2) \\
y &= a_1 \;\; \text{if } y^* \leq a_1 \\
y &= y^* \;\; \text{if } a_1 < y^* < a_2 \\
y &= a_2 \;\; \text{if } y^* \geq a_2
\end{aligned}
$$

• Not logically consistent if, say, $0 \leq y < 1$ because the Tobit implies a pile up at both endpoints. (In 821B we will cover other approaches to fractional responses.)

- Density of y is the same as y^* for values in (a_1, a_2), which as the $Normal(\mathbf{x}\boldsymbol{\beta}, \sigma^2)$ form.
- The endpoint probabilities are

$$
\begin{aligned}
P(y = a_1|\mathbf{x}) &= \Phi((a_1 - \mathbf{x}\boldsymbol{\beta})/\sigma) \\
P(y = a_2|\mathbf{x}) &= \Phi(-(a_2 - \mathbf{x}\boldsymbol{\beta})/\sigma).
\end{aligned}
$$

- The log likelihood for a random draw i is

$$
\begin{aligned}
\log[f(y_i|\mathbf{x}_i;\boldsymbol{\theta}) &= 1[y_i = a_1]\log[\Phi((a_1 - \mathbf{x}_i\boldsymbol{\beta})/\sigma)] \\
&+ 1[y_i = a_2]\log[\Phi(-(a_2 - \mathbf{x}_i\boldsymbol{\beta})/\sigma)] \\
&+ 1[a_1 < y_i < a_2]\log[(1/\sigma)\phi((y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma)].
\end{aligned}
$$

- Well behaved and standard asymptotic theory for MLE applies. Easy to obtain Wald and LR statistics for exclusion restrictions.
- Many econometrics packages that estimate the standard Tobit model also allow specifying any lower and upper limit. In Stata,

```
tobit y x1 x2 ... xk, ll(a1) ul(a2)
```

- Can compute several expectations to obtain magnitudes of effects. For example, the expectation conditional on not being at a limit point is

$$E(y|\mathbf{x}, a_1 < y < a_2) = \mathbf{x}\boldsymbol{\beta} + \frac{\sigma[\phi((a_1 - \mathbf{x}\boldsymbol{\beta})/\sigma) - \phi((a_1 - \mathbf{x}\boldsymbol{\beta})/\sigma)]}{[\Phi((a_2 - \mathbf{x}\boldsymbol{\beta})/\sigma) - \Phi((a_1 - \mathbf{x}\boldsymbol{\beta})/\sigma)]}$$

where the term after $\mathbf{x}\boldsymbol{\beta}$ is the extension of the inverse Mills ratio.

- The so-called unconditional expectation can be gotten from

$$\begin{aligned} E(y|\mathbf{x}) &= a_1 P(y = a_1|\mathbf{x}) + P(a_1 < y < a_2|\mathbf{x}) E(y|\mathbf{x}, a_1 < y < a_2) \\ + a_2 P(y &= a_2|\mathbf{x}) \\ &= a_1 \Phi((a_1 - \mathbf{x}\boldsymbol{\beta})/\sigma) + P(a_1 < y < a_2|\mathbf{x}) E(y|\mathbf{x}, a_1 < y < a_2) \\ &\quad + a_2 \Phi(-(a_2 - \mathbf{x}\boldsymbol{\beta})/\sigma). \end{aligned}$$

- These equations are a bit cumbersome to work with, but they do allow us to obtain predicted values for an vector $\mathbf{x}$, once we have obtained the MLES.

- As with the single corner at zero, the partial effect of a continuous variable x_j on $E(y|\mathbf{x})$ simplifies to a remarkable degree:

$$\frac{\partial E(y|\mathbf{x})}{\partial x_j} = [\Phi((a_2 - \mathbf{x}\boldsymbol{\beta})/\sigma) - \Phi((a_1 - \mathbf{x}\boldsymbol{\beta})/\sigma)]\beta_j.$$

- Can compute partial effects at specific values of $\mathbf{x}$, and average partial effects, especially easy to compute for continuous explanatory variables. For APEs we have

$$\left(N^{-1} \sum_{i=1}^{N} [\Phi((a_2 - \mathbf{x}_i\hat{\boldsymbol{\beta}})/\hat{\sigma}) - \Phi((a_1 - \mathbf{x}_i\hat{\boldsymbol{\beta}})/\hat{\sigma})] \right)\hat{\beta}_j,$$

where the scale factor is between zero and one.

- To determine how linear model estimates compare for estimating APEs, we should compare the OLS estimates for continuous variables directly to the scaled Tobit coefficients.
- If $\hat{\gamma}_j$ is the OLS slope estimate on the continuous variable x_j, from the regression y_i on $\mathbf{x}_i$ using all of the data, then $\hat{\gamma}_j$ is compared to the scaled Tobit coefficient. (This is most meaningful for continuous x_j, and might not be especially useful for discrete covariates.)
- APEs for binary variables should be obtained from the conditional mean equations, where we difference the two expected values at the two settings of the binary variable, and then average the differences.

• From the latent variable formulation of the model, should be clear how to adapt the Smith-Blundell and Chamberlain-Mundlak procedures. In the latter, still assume $c_i = \psi + \bar{\mathbf{x}}_i \boldsymbol{\xi} + r_i + u_{it}$ where

$$r_i | \mathbf{x}_i \sim Normal(0, \sigma_r^2)$$

Then we use a two-limit Tobit with explanatory variables $(1, \mathbf{x}_{it}, \bar{\mathbf{x}}_i)$ (and probably year dummies, and maybe even time constant variables $\mathbf{w}_i$).

• Again can use the the RE two-limit Tobit that assumes $\{u_{it}\}$ is serially uncorrelated with variance σ_u^2, or can use pooled two-limit Tobit.