Microeconometrics

Data censoring

Alex Armand
alex.armand@novasbe.pt

Lecture summary

1. Introduction
2. The Type I Tobit Model
3. Specification issues
4. Two-Limit Tobit models
5. Interval-censored outcomes

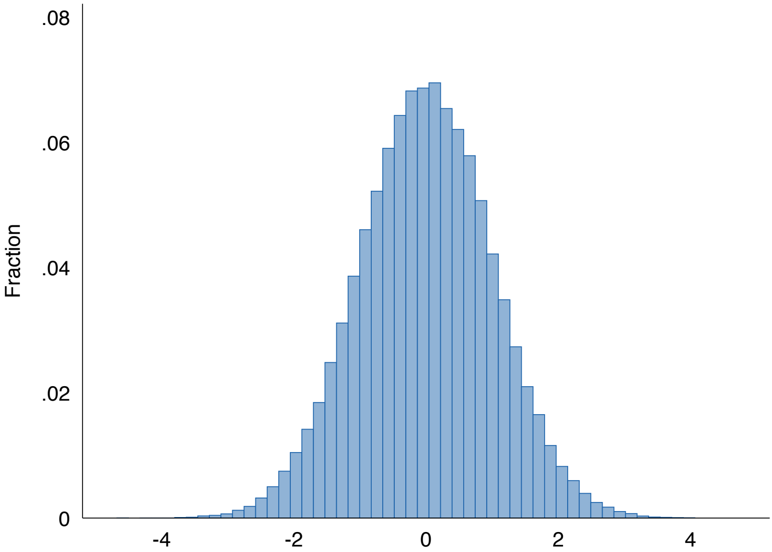
Censored versus truncated data

The common problem is **data observability**

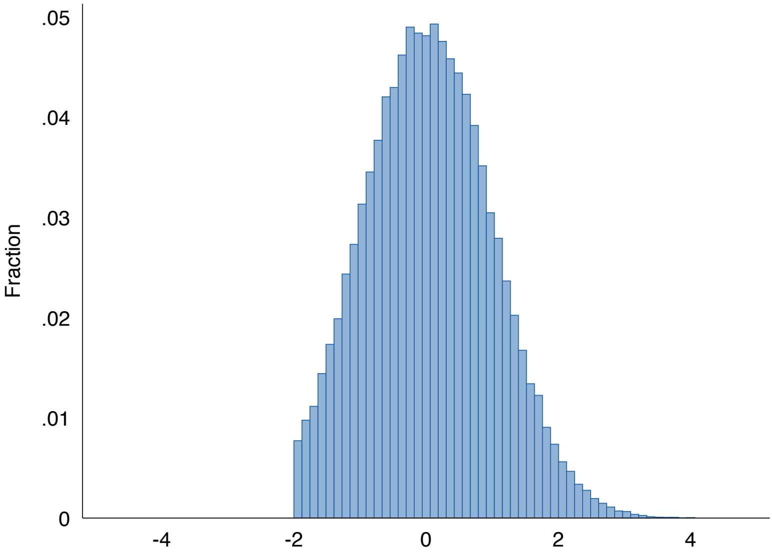
1. **Censored** data includes the censoring points
2. **Truncated** data excludes the censoring points

- Examples of data with this problem:
  - Earnings
  - Hours of work
  - Top coding of wealth
  - Expenditure on cars (Tobin's example)

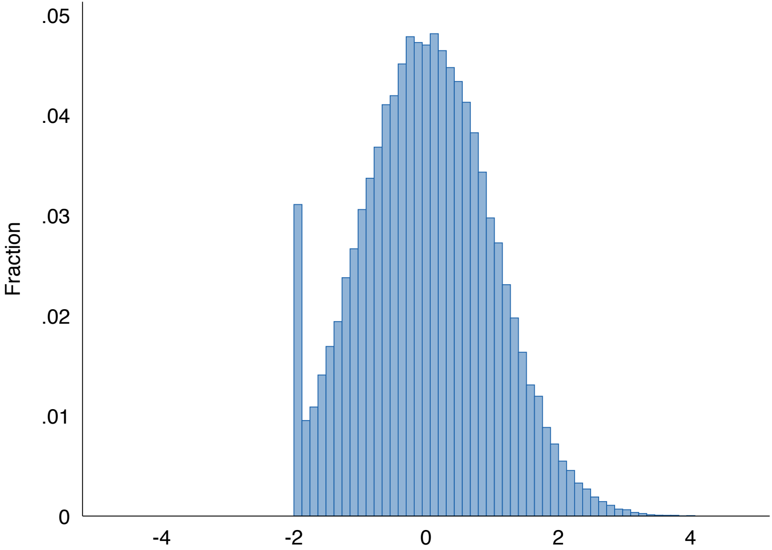
Example: Normal distribution (0,1)



Example: Normal distribution truncated

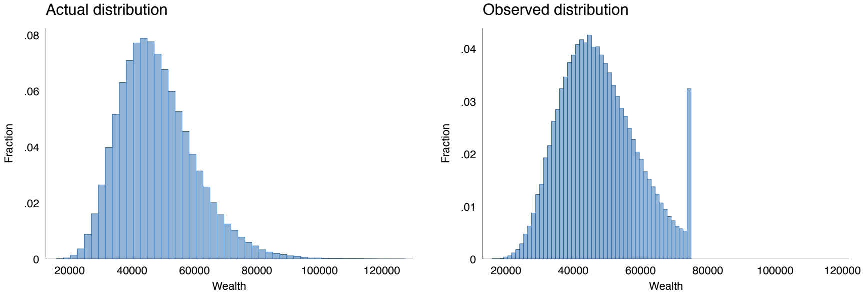


Example: Normal distribution censored



Example: top coding of wealth

- We are interested in measuring *wealth* for a certain population
  - $wealth^*$ denote *actual wealth* (a continuous variable)
  - *wealth* denote *observed wealth*
- Example: data are censored at 75k USD
  - This means that we observe $wealth = min(75, wealth^*)$



Censored regression models

Censored regression models are applied to two kinds of situations:

1. **Data censoring**
2. **Corner solution outcomes**: the variable we would like to explain piles up at one or two corners.

- Most empirical applications are actually to the second case
    - It happens to values at a corner (often zero).
    - Response is continuous over strictly positive values.
    - Examples:
        - Charitable contributions
        - Labor supply
        - Amount of life insurance

General formulation

Consider the case where $y \geq 0$ has a corner at zero.

- **Observability** of y can be written as

$$y = \max(0, \mathrm{x}\beta + u) \tag{1}$$

  - $\mathrm{x} = (1, x_2, ..., x_K)$
  - β is $K \times 1$
  - u is an unobserved error with some continuous distribution

- If the range of u is unrestricted
  - Equation (1) generates a pile up at zero and then continuous strictly positive outcomes

What can we say about $D(y|\mathrm{x})$ in general?

We want to know about the full distribution of y, not only the observable part

- We need to restrict $D(u|\mathrm{x})$ in some way.
- **Example**:

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

  - We can pass the median through:

$$\begin{aligned} Med(y|\mathrm{x}) &= Med[\max(0, \mathrm{x}\beta + u)|\mathrm{x}] \\ &= \max[0, Med(\mathrm{x}\beta + u|\mathrm{x})] \\ &= \max[0, \mathrm{x}\beta + Med(u|\mathrm{x})] \\ &= \max(0, \mathrm{x}\beta) \end{aligned}$$

What about $E(y|\mathrm{x})$?

Generally, we cannot find $E(y|\mathrm{x})$ without much stronger assumptions.

- Function $\max(0, z)$ is a *convex function* $\Rightarrow$ the line segment between any two distinct points on the graph of the function lies above the graph between the two points.
- *Convex function* $\Rightarrow$ **Jensen's inequality**

$$E(y|\mathrm{x}) \equiv E[\max(0, \mathrm{x}\beta + u)|\mathrm{x}] \geq \max[0, \mathrm{x}\beta + E(u|\mathrm{x})] \equiv \max(0, \mathrm{x}\beta)$$

- We can only get a **lower bound** for $E(y|\mathrm{x})$

$$E(y|\mathrm{x}) \geq \max(0, \mathrm{x}\beta) = med(y|\mathrm{x})$$

- This is not sufficient, if we want to learn more about $E(y|\mathrm{x})$ we need more assumption $\rightarrow$ **Tobit model**

Tobit Model

By far the **most popular model for corners at zero**

- **Type I**: assumes **censoring at zero** and normality + homoskedasticity of the error term

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

- Similarly to probit and logit we need to assume a distribution for the error term to derive a density for MLE
- Can be seen as a latent variable model for y^*
  - $y^* = \mathrm{x}\beta + u$
  - $D(y^*|\mathrm{x})$ follows a classical linear model

Quantities of interest and observability

Our objective is to learn about **partial effects** $\frac{\partial E(y|\mathrm{x})}{\partial x_j}$

- $E(y|\mathrm{x})$ is not observed because we observe y only if positive
- Notice that we can write it as

$$E(y|\mathrm{x}) = P(y > 0|\mathrm{x}) \cdot E(y|\mathrm{x}, y > 0) + P(y = 0|\mathrm{x}) \cdot E(y|\mathrm{x}, y = 0)$$

- **Can we recover these components separately?**
  1. $P(y > 0|\mathrm{x})$: probability of y being observed
  2. $E(y|\mathrm{x}, y > 0)$: conditional mean for observable data
  3. $P(y = 0|\mathrm{x})$
  4. $E(y|\mathrm{x}, y = 0)$: notice this is equal to zero!

Quantities of interest and observability

Our objective is to learn about **partial effects** $\frac{\partial E(y|\mathrm{x})}{\partial x_j}$

- $E(y|\mathrm{x})$ is not observed because we observe y only if positive
- Notice that we can write it as

$$E(y|\mathrm{x}) = P(y > 0|\mathrm{x}) \cdot E(y|\mathrm{x}, y > 0) + P(y = 0|\mathrm{x}) \cdot E(y|\mathrm{x}, y = 0)$$

- **Can we recover these components separately?**
  1. $P(y > 0|\mathrm{x})$: probability of y being observed
  2. $E(y|\mathrm{x}, y > 0)$: conditional mean for observable data
  3. $P(y = 0|\mathrm{x})$
  4. $E(y|\mathrm{x}, y = 0)$: notice this is equal to zero!

Recovering $P(y > 0|\mathrm{x})$

- Compute response probability using $u|\mathrm{x} \sim \mathcal{N}(0, \sigma^2)$:

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

- It follows a **probit model** with parameter vector β/σ
- We already know how to estimate it and obtain partial effects

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

Recovering $E(y|\mathrm{x}, y > 0)$

- Revision from statistics: if $z \sim \mathcal{N}(0,1)$ then

$$E(z|z > c) = \phi(c)/[1 - \Phi(c)]$$

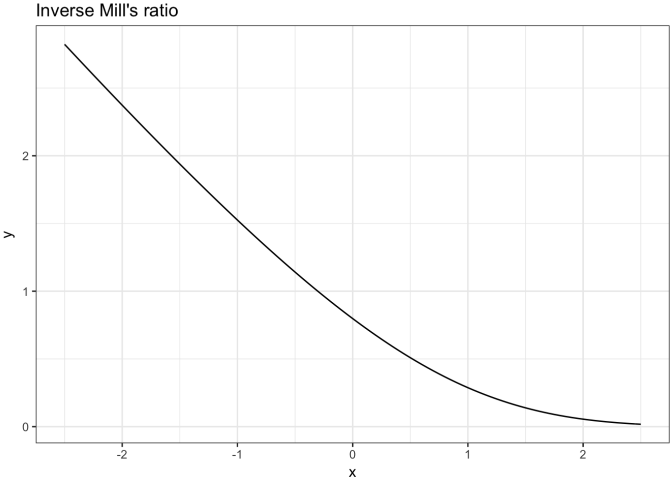
- We can write the **conditional mean for observable data**

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

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

Inverse Mills ratio

$$\lim_{z\to\infty} \lambda(z) = 0 \qquad \lim_{z\to-\infty} \lambda(z) = \infty$$



Unconditional expectation $E(y|\mathrm{x})$

Since we derived all components, we can derive $E(y|\mathrm{x})$

- The **unconditional expectation** is equal to

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

- Called the unconditional expectation, since we are not conditioning on $y > 0$, even though we condition on x

Partial effects for conditional expectations

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

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

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

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

- If x_j and x_h are two continuous variables, the ratio of partial effects is β_j/β_h

Partial effects for unconditional expectations

For the unconditional expectation, a generally useful expression is

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

- Applied to the **Type 1 Tobit model**

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

- Again β_j is **scaled by a function between** 0 **and** 1 which depends on x
  - As $P(y>0|\mathrm{x}) \to 1$ the β_j become close to the actual partial effect.
  - If $P(y=0|\mathrm{x})$ is large, the scale factor is small

Estimation of parameters

Suppose we have a random sample from the population

$$\{(\mathrm{x}_i, y_i) : i = 1, 2, ..., N\}$$

- OLS regressions y_i on x_i using the full sample or the sample with $y_i > 0$ does not consistently estimate $\beta \rightarrow$ use MLE

1. Density function for **zero-values**

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

2. Density function for **positive values**: for $y > 0$, $f(y|\mathrm{x}) = f^*(y|\mathrm{x})$, where $y|\mathrm{x} \sim \mathcal{N}(\mathrm{x}\beta, \sigma^2)$

$$\begin{aligned} f(y|\mathrm{x}, y > 0) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \mathrm{x}\beta)^2}{2\sigma^2}} \\ &= \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y_i - \mathrm{x}\beta}{\sigma})} = \frac{1}{\sigma}\phi(\frac{y - \mathrm{x}\beta}{\sigma}) \end{aligned}$$

MLE procedure

1. Write the **density** of y

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

where $1[condition]$ is 1 if the condition is true, and 0 otherwise

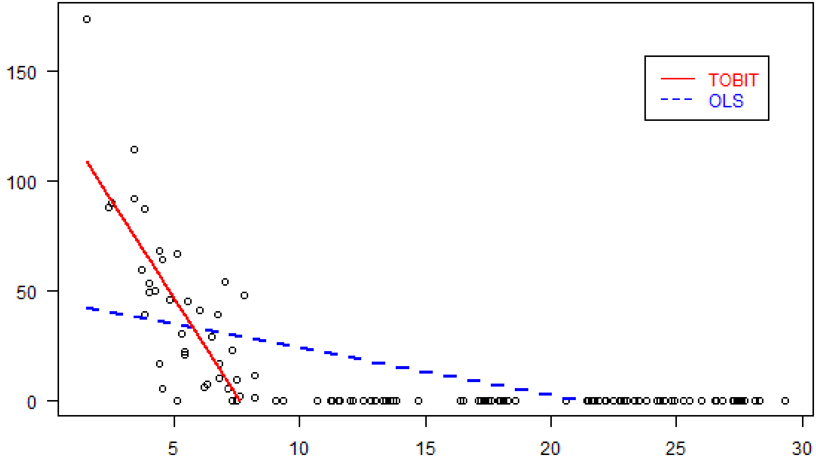
2. Write the **log-likelihood** for random draw i

$$\begin{aligned} \ell_i(\beta, \sigma) &= 1[y_i = 0] \log[1 - \Phi(\mathrm{x}_i\beta/\sigma)] \\ +1[y_i &> 0]\{\log \phi[(y_i - \mathrm{x}\beta)/\sigma] - \log(\sigma)\} \end{aligned}$$

3. Compute optimality conditions to find MLE estimates

Interpretation of Tobit

Compare the Tobit APEs on $E(y|\mathrm{x})$ to OLS estimates using entire sample



Goodness-of-fit

Not a single straightforward procedure

1. If we focus on $E(y|\mathrm{x})$, a simple metric is the squared correlation between y_i and $\hat{E}(y_i|\mathrm{x}_i)$

$$\hat{E}(y_i|\mathrm{x}_i) = \Phi(\mathrm{x}_i\hat{\beta}/\hat{\sigma})\mathrm{x}_i\hat{\beta} + \hat{\sigma}\phi(\mathrm{x}_i\hat{\beta}/\hat{\sigma}).$$

2. We can use a sum of squared residuals-type R-squared, comparable to OLS R-squared.
3. We can look at the fit for nonlimit observations

$$\hat{E}(y_i|\mathrm{x}_i, y_i > 0) = \mathrm{x}_i\hat{\beta} + \hat{\sigma}\lambda(\mathrm{x}_i\hat{\beta}/\hat{\sigma})$$

Example: Married Labor Force Participation (Mroz 1987)

THE SENSITIVITY OF AN EMPIRICAL MODEL OF MARRIED WOMEN'S HOURS OF WORK TO ECONOMIC AND STATISTICAL ASSUMPTIONS

BY THOMAS A. MROZ[1]

This study undertakes a systematic analysis of several theoretic and statistical assumptions used in many empirical models of female labor supply. Using a single data set (PSID 1975 labor supply data) we are able to replicate most of the range of estimated income and substitution effects found in previous studies in this field. We undertake extensive specification tests and find that most of this range should be rejected due to statistical and model misspecifications. The two most important assumptions appear to be (i) the Tobit assumption used to control for self-selection into the labor force and (ii) exogeneity assumptions on the wife's wage rate and her labor market experience. The Tobit models exaggerate both the income and wage effects. The exogeneity assumptions induce an upwards bias in the estimated wage effect; the bias due to the exogeneity assumption on the wife's labor market experience, however, substantially diminishes when one controls for self-selection into the labor force through the use of unrestricted generalized Tobit procedures. An examination of the maintained assumptions in previous studies further supports these results. These inferences suggest that the small responses to variations in wage rates and nonwife income found here provide a more accurate description of the behavioral responses of working married women than those found in most previous studies.

Dataset

```
. des nwifeinc educ exper expersq age kidslt6 kidsge6

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------
nwifeinc        float  %9.0g                  (faminc - wage*hours)/1000
educ            byte   %9.0g                  years of schooling
exper           byte   %9.0g                  actual labor mkt exper
expersq         int    %9.0g                  exper^2
age             byte   %9.0g                  woman's age in yrs
kidslt6         byte   %9.0g                  # kids < 6 years
kidsge6         byte   %9.0g                  # kids 6-18

. sum nwifeinc educ exper expersq age kidslt6 kidsge6

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    nwifeinc |       753    20.12896     11.6348  -.0290575         96
        educ |       753    12.28685    2.280246          5         17
       exper |       753    10.63081     8.06913          0         45
     expersq |       753    178.0385    249.6308          0       2025
         age |       753    42.53785    8.072574         30         60
-------------+--------------------------------------------------------
     kidslt6 |       753    .2377158     .523959          0          3
     kidsge6 |       753    1.353254    1.319874          0          8
```

Probit

```
. probit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

Probit regression                                 Number of obs   =        753
                                                  LR chi2(7)      =     227.14
                                                  Prob > chi2     =     0.0000
Log likelihood = -401.30219                       Pseudo R2       =     0.2206

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0120237   .0048398   -2.48   0.013    -.0215096   -.0025378
        educ |   .1309047   .0252542    5.18   0.000     .0814074     .180402
       exper |   .1233476   .0187164    6.59   0.000     .0866641    .1600311
     expersq |  -.0018871      .0006   -3.15   0.002     -.003063   -.0007111
         age |  -.0528527   .0084772   -6.23   0.000    -.0694678   -.0362376
     kidslt6 |  -.8683285   .1185223   -7.33   0.000    -1.100628    -.636029
     kidsge6 |    .036005   .0434768    0.83   0.408     -.049208    .1212179
       _cons |   .2700768    .508593    0.53   0.595    -.7267473    1.266901
------------------------------------------------------------------------------
```

APE after Probit

```
. margeff

Average partial effects after probit
      y  = Pr(inlf)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0036162   .0014414    -2.51   0.012    -.0064413   -.0007911
        educ |   .0393088   .0071877     5.47   0.000     .0252212    .0533964
       exper |    .037046    .005131     7.22   0.000     .0269893    .0471026
     expersq |  -.0005675   .0001771    -3.20   0.001    -.0009146   -.0002204
         age |  -.0158917   .0023569    -6.74   0.000     -.020511   -.0112723
     kidslt6 |  -.2441788   .0258995    -9.43   0.000    -.2949409   -.1934167
     kidsge6 |   .0108274   .0130538     0.83   0.407    -.0147576    .0364124
------------------------------------------------------------------------------
```

Number of hours worked

```
. sum hours

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


. count if hours == 0
  325
```

OLS estimates on full sample

```
. 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
------------------------------------------------------------------------------
```

Tobit estimates

```
. 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
```

APE after tobit

- Notice that tobit command in STATA is already reporting

```
. margins, dydx(*)

Average marginal effects                        Number of obs     =        753
Model VCE    : OIM

Expression   : Linear prediction, predict()
dy/dx w.r.t. : nwifeinc educ exper expersq age kidslt6 kidsge6

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -8.814226   4.459089    -1.98   0.048    -17.56808   -.0603706
        educ |   80.64541   21.58318     3.74   0.000     38.27441    123.0164
       exper |    131.564   17.27935     7.61   0.000     97.64211     165.486
     expersq |  -1.864153   .5376606    -3.47   0.001    -2.919661   -.8086455
         age |  -54.40491   7.418483    -7.33   0.000     -68.9685   -39.84133
     kidslt6 |  -894.0202   111.8777    -7.99   0.000    -1113.653   -674.3875
     kidsge6 |  -16.21805    38.6413    -0.42   0.675    -92.07668    59.64057
------------------------------------------------------------------------------
```

APE on $E[y|x, y > 0]$ after tobit

- Closer to OLS estimates

```
. margins, dydx(*) predict (ystar(0,.))

Average marginal effects                          Number of obs     =        753
Model VCE    : OIM

Expression   : E(hours*|hours>0), predict(ystar(0,.))
dy/dx w.r.t. : nwifeinc educ exper expersq age kidslt6 kidsge6

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -5.188619   2.621409    -1.98   0.048    -10.32649   -.0507514
        educ |   47.47306    12.6214     3.76   0.000     22.73558    72.21054
       exper |   77.44703    9.99765     7.75   0.000     57.85199    97.04206
     expersq |   -1.09736   .3155945    -3.48   0.001    -1.715914   -.4788063
         age |  -32.02622    4.29211    -7.46   0.000     -40.4386   -23.61384
     kidslt6 |  -526.2776   64.70619    -8.13   0.000    -653.0994   -399.4558
     kidsge6 |  -9.546986   22.75224    -0.42   0.675    -54.14056    35.04659
------------------------------------------------------------------------------
```

Goodness-of-fit

Compute the squared correlation between y_i and $\hat{E}(y_i|\mathrm{x}_i)$

- Predict the number of hours and then compute the squared correlation

```
. 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
```

Visualizing results: Tobit

- Assume for simplicity we want to focus on hours and nwifeinc only (no other controls)

```
. tobit hours nwifeinc, ll(0)

Refining starting values:

Grid node 0:   log likelihood = -4070.7304

Fitting full model:

Iteration 0:   log likelihood = -4070.7304
Iteration 1:   log likelihood = -3962.7553
Iteration 2:   log likelihood = -3948.8929
Iteration 3:   log likelihood = -3948.1309
Iteration 4:   log likelihood = -3948.1289
Iteration 5:   log likelihood = -3948.1289

Tobit regression                                Number of obs     =        753
                                                   Uncensored     =        428
Limits: lower = 0                                  Left-censored  =        325
        upper = +inf                               Right-censored =          0

                                                LR chi2(1)        =      13.53
                                                Prob > chi2       =     0.0002
Log likelihood = -3948.1289                     Pseudo R2         =     0.0017

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -17.57587   4.835457    -3.63   0.000    -27.06847   -8.083274
       _cons |   665.8369   109.6933     6.07   0.000     450.4954    881.1784
-------------+----------------------------------------------------------------
 var(e.hours)|    1853209   140396.3                       1597110     2150375
------------------------------------------------------------------------------
```

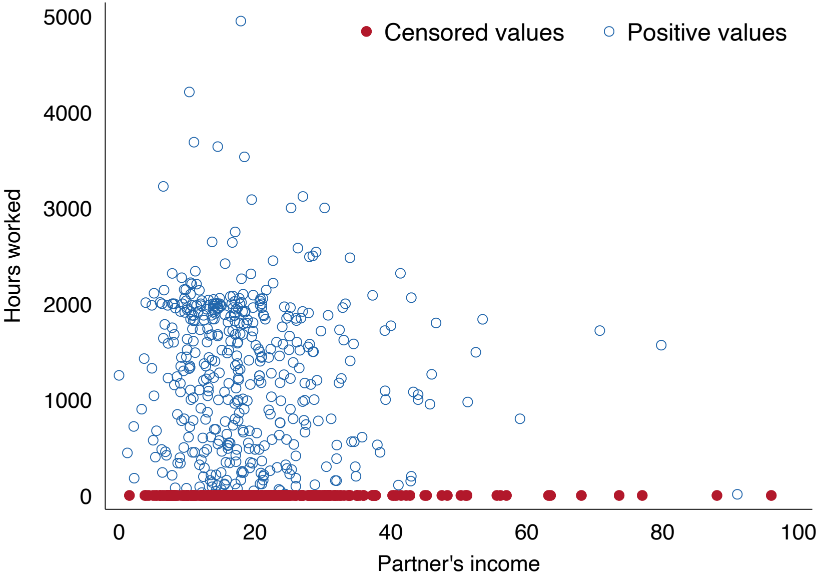
Visualizing results: OLS on full sample

```
. reg hours nwifeinc

      Source |       SS           df       MS      Number of obs   =       753
-------------+----------------------------------   F(1, 751)       =     11.86
       Model |  8873133.93         1  8873133.93   Prob > F        =    0.0006
    Residual |   562036590       751  748384.274   R-squared       =    0.0155
-------------+----------------------------------   Adj R-squared   =    0.0142
       Total |   570909724       752  759188.463   Root MSE        =    865.09

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -9.336214   2.711406    -3.44   0.001    -14.65905   -4.013377
       _cons |   928.5047   63.02861    14.73   0.000     804.7714    1052.238
------------------------------------------------------------------------------
```

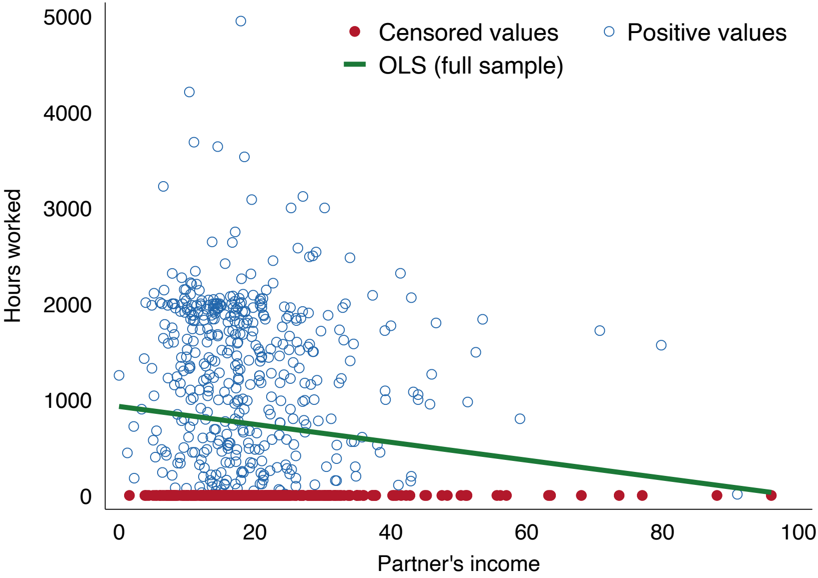
Visualizing results: OLS on positive values only

```
. reg hours nwifeinc if hours > 0

      Source |       SS           df       MS      Number of obs   =       428
-------------+----------------------------------   F(1, 426)       =      2.85
       Model |  1707248.02         1  1707248.02   Prob > F        =    0.0924
    Residual |   255603772       426  600008.854   R-squared       =    0.0066
-------------+----------------------------------   Adj R-squared   =    0.0043
       Total |   257311020       427   602601.92   Root MSE        =     774.6

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc | -5.970117   3.539268    -1.69   0.092    -12.92672    .9864838
       _cons |  1415.989    76.7738    18.44   0.000     1265.086    1566.892
------------------------------------------------------------------------------
```

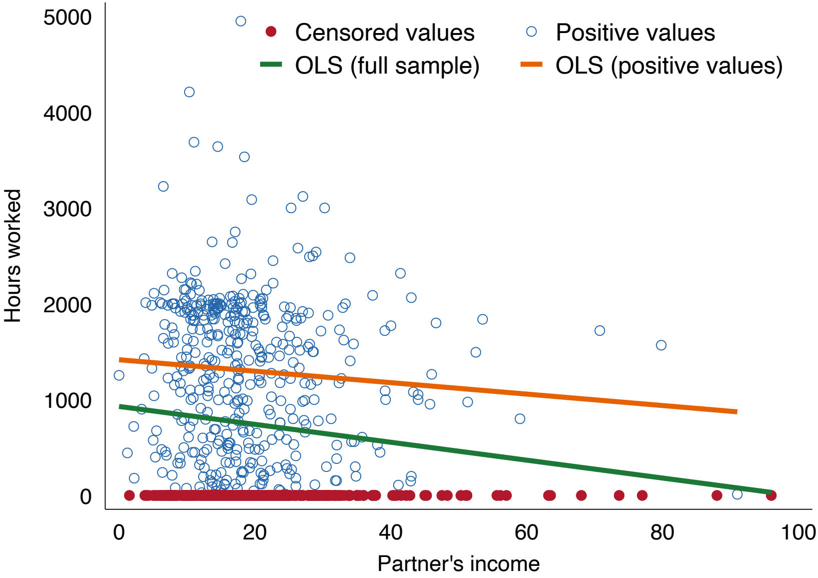
Visualizing results: plot the results



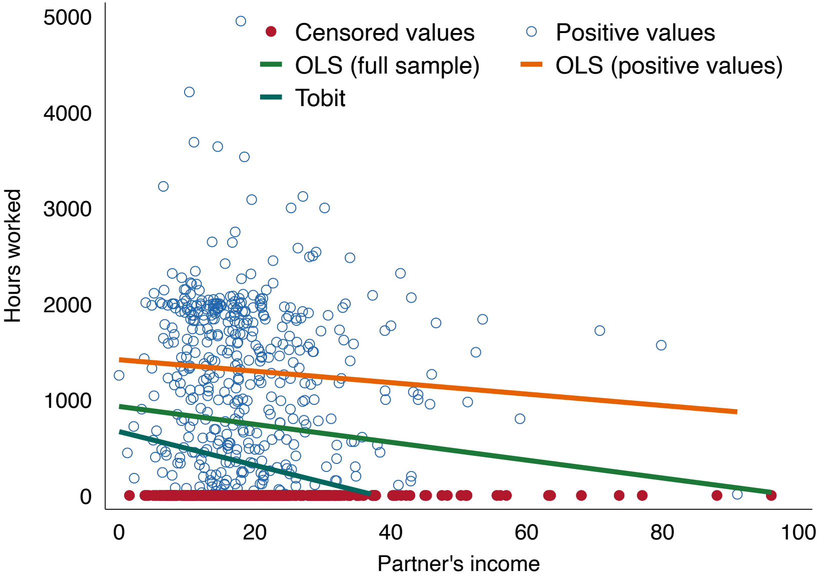
Visualizing results: plot the results



Visualizing results: plot the results



Visualizing results: plot the results



Specification issues in tobit models

- Let's look at some cases using the tobit model
  - **Omitted heterogeneity independent of the covariates**
  - **Heteroskedasticity**
  - **Non-normality**

Omitted heterogeneity independent of the covariates

Conclusions are similar to the binary response case

- Assume we have q as unobserved heterogeneity

$$\begin{aligned} y &= \max(0, \mathrm{x}\beta + \gamma q + u) \\ u|(\mathrm{x}, q) &\sim \mathcal{N}(0, \sigma^2) \\ q|\mathrm{x} &\sim \mathcal{N}(0, \tau^2) \end{aligned}$$

- If we estimate a standard tobit we are instead assuming

$$\begin{aligned} y &= \max(0, \mathrm{x}\beta + v) \\ v|(\mathrm{x}) &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

- But notice that v is not distributed $\mathcal{N}(0, \sigma^2)$

Specification issues in tobit models

- Let's look at some cases using the tobit model
  - **Omitted heterogeneity independent of the covariates**
  - **Heteroskedasticity**
  - **Non-normality**

Heteroskedasticity

- Again similar to result for probit
- **Heteroskedastic Tobit** is a good way to extend functional form.
- Typically we can assume that

$$u|\mathrm{x} \sim \mathcal{N}(0, \exp(2\mathrm{x}\delta))$$

- Similar to probit, it makes the partial effects on $E(y|\mathrm{x}, y > 0)$ and $E(y|\mathrm{x})$ more difficult to estimate.

Specification issues in tobit models

- Let's look at some cases using the tobit model
  - **Omitted heterogeneity independent of the covariates**
  - **Heteroskedasticity**
  - **Non-normality**

Non-normality

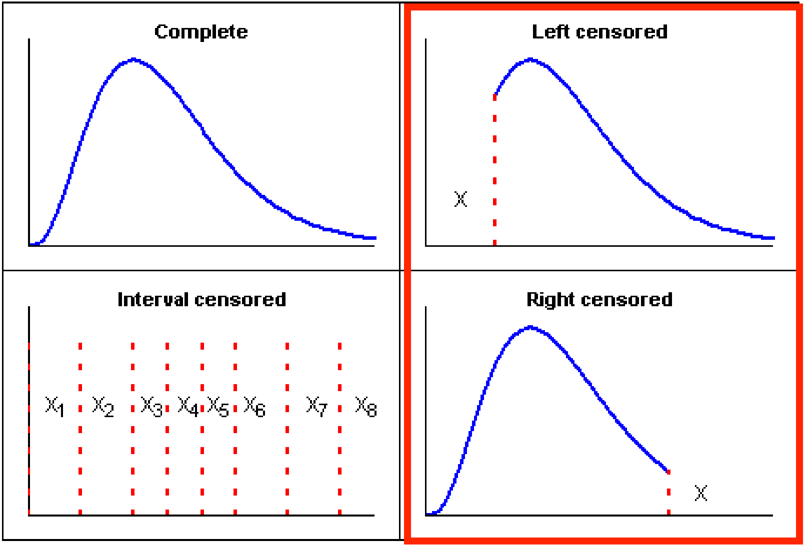
Assume the following model

$$\begin{aligned} y &= \max(0, \mathrm{x}\beta + u) \\ u|(\mathrm{x}, q) &\sim \mathcal{F}(\cdot) \end{aligned}$$

- The usual Tobit MLE will not consistently estimate β
  - It may yield reasonably close partial effects
  - Using a more flexible distribution for $D(u|\mathrm{x})$ might be a good idea, but one should not only compare estimated coefficients.

Extention I: two-limit tobit models

- Focus now on cases in which censoring is at multiple points



Extention I: two-limit tobit models

- Allow for **two censoring points**
  - These might be logical or institutional constraints.
  - Common are corners at 0 and 1 or 0 and 100.
- Example:
  - 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
    - What would happen if the cap were not there?
    - What would happen if it is raised?

Extention I: two-limit tobit models

- Let $a_1 < a_2$ be the two limit values of y in the population

$$
\begin{aligned}
y^* &= \mathrm{x}\beta + u, \quad u|\mathrm{x} \sim \mathcal{N}(0, \sigma^2) \\
y &= a_1 \ \text{ if } y^* \leq a_1 \\
y &= y^* \ \text{ if } a_1 < y^* < a_2 \\
y &= a_2 \ \text{ if } y^* \geq a_2
\end{aligned}
$$

- **Endpoint probabilities** are

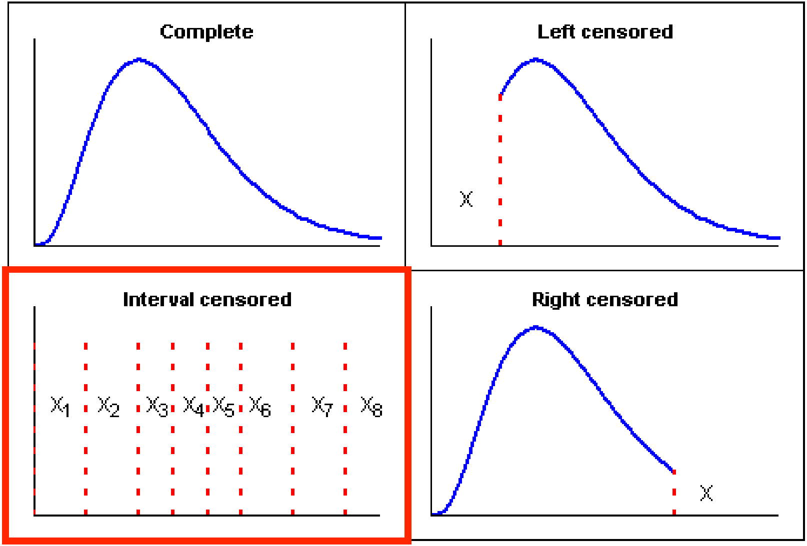
$$
\begin{aligned}
P(y &= a_1|\mathrm{x}) = \Phi((a_1 - \mathrm{x}\beta)/\sigma) \\
P(y &= a_2|\mathrm{x}) = \Phi(-(a_2 - \mathrm{x}\beta)/\sigma).
\end{aligned}
$$

- **Log-likelihood for a random draw** i is

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

Extension II: interval-coded date

- Focus now on cases in which data is interval-coded



Extension II: interval-coded date

- **Interval-coded data** or **interval-censored data**: the response variable is recorded in intervals, but the underlying variable is continuous
  - For example, rather than asking individuals to report actual annual income, they report the interval that their income falls into.
- Let $r_1 < r_2 < ... < r_J$ denote the *known* interval limits

$$
\begin{aligned}
w &= 0 \quad \text{if } y \leq r_1 \\
w &= 1 \quad \text{if } r_1 < y \leq r_2 \\
&\vdots \\
w &= J \quad \text{if } y > r_J
\end{aligned}
$$

Extension II: interval-coded date

- Expand the tobit model to have many censoring points

- The **log-likelihood** for a random draw i is

$$\begin{aligned}\ell_i(\beta,\sigma) &= 1[w_i = 0]\log\{\Phi[(r_1 - x_i\beta)/\sigma]\} \\ &+1[w_i = 1]\log\{\Phi[(r_2 - x_i\beta)/\sigma] - \Phi[(r_1 - x_i\beta)/\sigma]\} \\ &... + 1[w_i = J]\log\{1 - \Phi[(r_J - x_i\beta)/\sigma]\}\end{aligned}$$

- The MLE, $\hat{\beta}$ and $\hat{\sigma}^2$, are often called **interval regression** estimators