Lecture 8
Models for Censored and Truncated Data - Tobit Model

1

(For private use, not to be posted/shared online).

Censored and Truncated Data

• In some data sets we do not observe values above or below a certain magnitude, due to a censoring or truncation mechanism.

Examples:

- A central bank intervenes to stop an exchange rate falling below or going above certain levels.
- Dividends paid by a company may remain zero until earnings reach some threshold value.
- A government imposes price controls on some goods.
- A survey of only working women, ignoring non-working women.

In these situations, the observed data consists of a combination of measurements of some underlying *latent variable* and observations that arise when the censoring/truncation mechanism is applied. 2

Censored and Truncated Data: Definitions

• Y is **censored** when we observe X for all observations, but we only know the true value of Y for a restricted range of observations. Values of Y in a certain range are reported as a single value or there is significant clustering around a value, say 0.

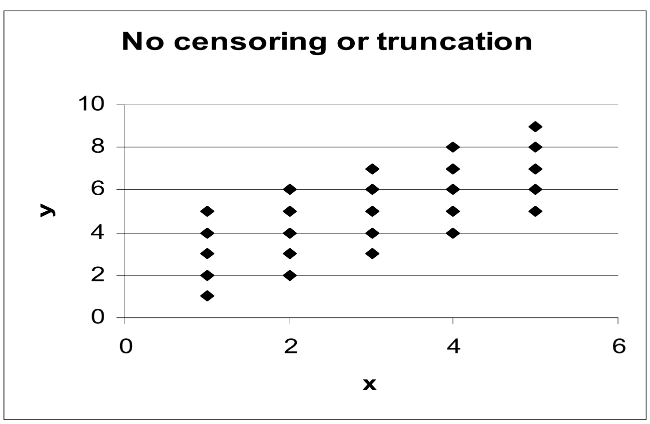
- If $Y = k$ or $Y > k$ for all $Y \Rightarrow Y$ is *censored from below* or *left-censored.*

- If $Y = k$ or $Y < k$ for all $Y \Rightarrow Y$ is *censored from above* or *right-censored.*

We usually think of an uncensored Y, Y^*, the true value of Y when the censoring mechanism is not applied. We typically have all the observations for $\{Y,X\}$, but not $\{Y^*, X\}$.

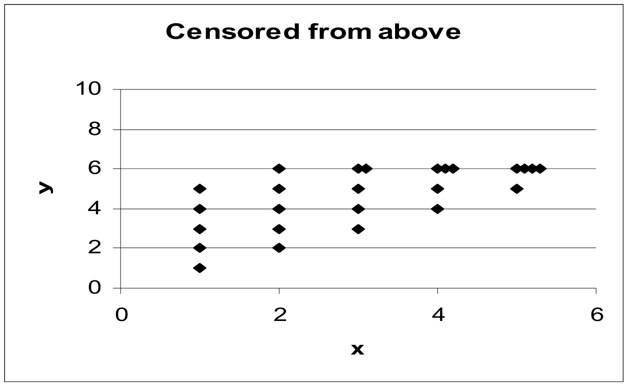
• Y is **truncated** when we only observe X for observations where Y would not be censored. We do not have a full sample for $\{Y, X\}$, we exclude observations based on characteristics of Y.

Censored and Truncated Data: Example 1



• We observe the full range of Y and the full range of X.

Censored and Truncated Data: Example 2

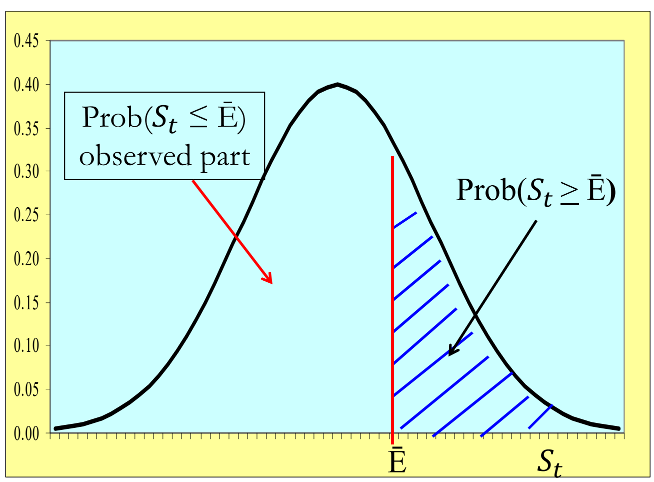


• If $Y \geq 6$, we do not know its exact value.

Example: A Central Bank intervenes if the exchange rate, S_t, hits the band's upper limit. Thus, if $S_t \geq \bar{E}$ $\Rightarrow S_t = \bar{E}$

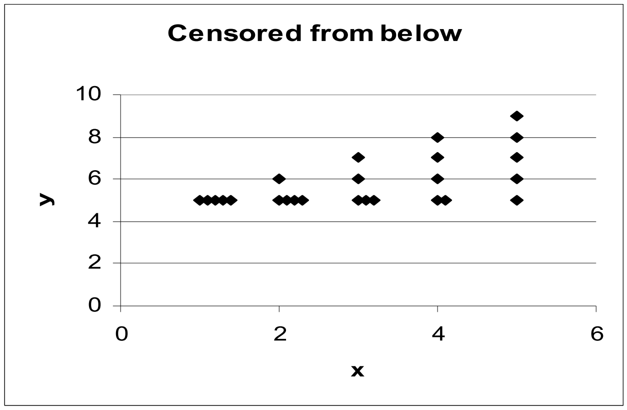
Censored and Truncated Data: Example 2

Pdf(S_t)



• The pdf of the exchange rate, S_t, is a mixture of discrete (mass at $S_t = \bar{E}$) and continuous (Prob[$S_t < \bar{E}$]) distributions.

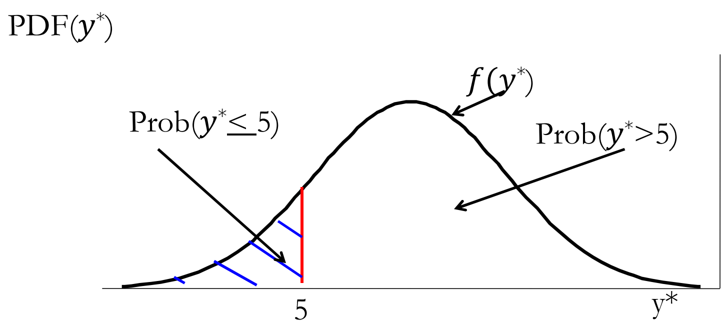
Censored and Truncated Data: Example 3



• If $Y \leq 5$, we do not know its exact value.

Example: A Central Bank intervenes if the exchange rate hits the band's lower limit. Thus, if $S_t \leq \bar{E}$ $\Rightarrow S_t = \bar{E}$

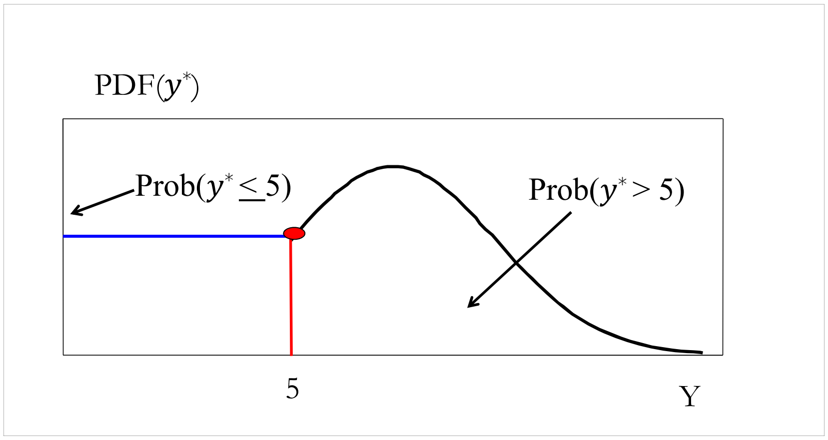
Censored and Truncated Data: Example 3



• The pdf of the observable variable, y, is a mixture of discrete (prob. mass at Y=5) and continuous (Prob[y^*>5]) distributions.

8

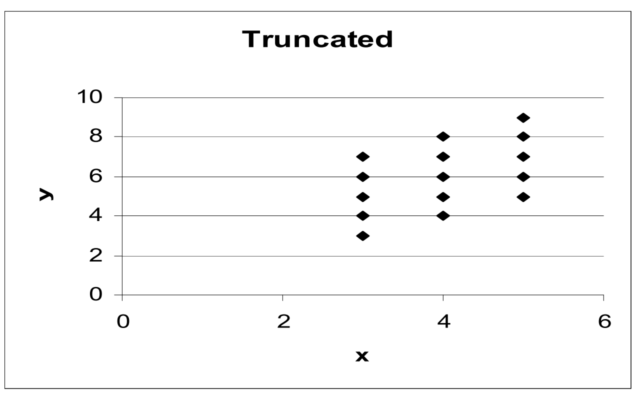
Censored and Truncated Data: Example 3



• Under censoring we assign the full probability in the censored region to the censoring point, 5.

9

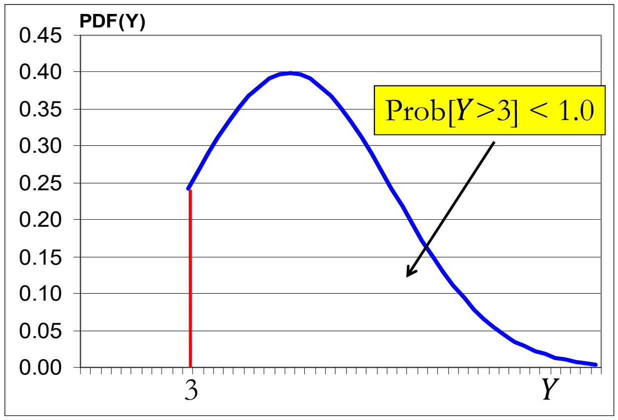
Censored and Truncated Data: Example 4



• If $Y < 3$, the value of X (or Y) is unknown. *(Truncation from below.)*

Example: If a family's income is below certain level, we have no information about the family's characteristics.

Censored and Truncated Data: Example 4



• Under data censoring, the censored distribution is a combination of a pmf plus a pdf. They add up to 1. We have a different situation under truncation. To create a pdf for Y we will use a conditional pdf.

Censored Normal

• Moments:

Let $y^* \sim N(\mu^*, \sigma^2)$ and $\alpha = \frac{(c - \mu^*)}{\sigma}$. Then

- Data: $y_i = y_i^*$ if $y_i^* \geq c$
  $= c$ if $y_i^* \leq c$

- $\text{Prob}(y = c \mid x) = \text{Prob}(y^* \leq c \mid x) = \text{Prob}[\frac{(y^* - \mu^*)}{\sigma} \leq \frac{(c - \mu^*)}{\sigma} \mid x]$
  $= \text{Prob}[z \leq \frac{(c - \mu^*)}{\sigma} \mid x] = \Phi(\alpha)$

- $\text{Prob}(y > c \mid x) = \text{Prob}(y^* > c \mid x) = 1 - \Phi(\alpha)$

- First Moment
  - Conditional: $E[y \mid y^* > c] = \mu^* + \sigma\,\lambda(\alpha)$
  - Unconditional: $E[y] = \Phi(\alpha)\, c + (1 - \Phi(\alpha))\, [\mu^* + \sigma\,\lambda(\alpha)]$

12

Censored Normal

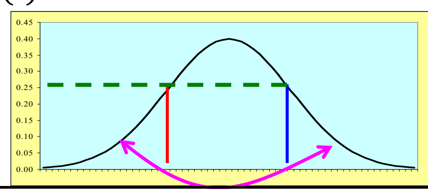
• To get the first moment, we use a useful result –proven later- for a truncated normal distribution. If v is a standard normal variable and the truncation is from below at c, a constant, then

$$\text{E}[v \mid v > c) = F^{-1}f = \frac{\phi\ (c)}{1-\Phi(c)} = \frac{\phi\ (-c)}{\Phi(-c)}$$

In our conditional model, c = -($\boldsymbol{x}_i{}'\beta$). (Note that the expectation is also conditioned on $\boldsymbol{x}$, thus $\boldsymbol{x}$ is treated as a constant.).

Note: The ratio $F^{-1}f$ (a pdf divided by a CDF) is called ***Inverse Mill's ratio***, usually denoted by $\lambda(.)$ –the ***hazard function***. If the truncation is from above:

$$\lambda(c) = \frac{\phi(c)}{1-\Phi(c)}$$



Censored Normal

• Moments (continuation):

- **Unconditional first moment**:

$$\text{E}[y] = \Phi(\alpha)\, c + (1 - \Phi(\alpha))\, [\mu^* + \sigma\, \lambda(\alpha)]$$

If $c = 0 \Rightarrow$ $\quad \text{E}[y] = (1 - \Phi(\alpha))\, [\mu^* + \sigma\, \lambda(\alpha)]$ $\quad (\alpha = \frac{(c - \mu^*)}{\sigma})$

$$= \Phi(\frac{\mu^*}{\sigma})\, [\mu^* + \sigma\, \lambda(\alpha)]$$

- **Second Moment**

$$\text{Var}[y^* \mid y > c] = \sigma^2 * [1 - \Phi(\alpha)] * [(1 - \delta) + (\lambda - \alpha)^2\, \Phi(\alpha)]$$

where $\delta = \lambda * [\lambda - \alpha]$ $\qquad 0 \leq \delta \leq 1$

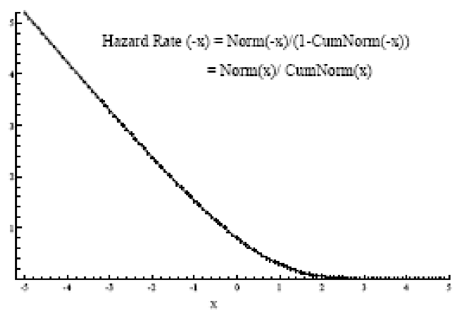
14

Censored Normal – Hazard Function

• The moments depend on $\lambda(c)$, the Inverse Mill's ratio or hazard rate evaluated at c:

$$\lambda(c) = \frac{\phi(c)}{1 - \Phi(c)} = \frac{\phi(-c)}{\Phi(-c)}$$

It is a monotonic function that begins at zero (when $c = -\infty$) and asymptotes at infinity (when $c = \infty$). See plot below for $(-c)$.



15

Truncated Normal

• Moments:

Let $y^* \sim N(\mu^*, \sigma^2)$ and $\alpha = \frac{(c - \mu^*)}{\sigma}$.

- **First moment**:

$E[y^* | y > c] = \mu^* + \sigma\, \lambda(\alpha)$ <= This is the truncated regression.

⇒ If μ >0 and the truncation is from below –i.e., $\lambda(\alpha)$>0–, the mean of the truncated variable is greater than the original mean

Note: For the standard normal distribution $\lambda(\alpha)$ is the mean of the truncated distribution.

- **Second moment**:

- $Var[y^* | y > c] = \sigma^2[1 - \delta(\alpha)]$ where $\delta(\alpha) = \lambda(\alpha)\, [\lambda(\alpha) - \alpha]$

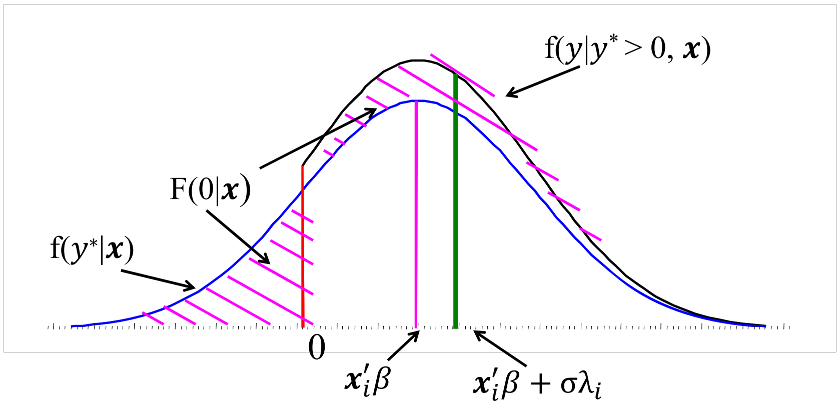
⇒ Truncation reduces variance! This result is general, it applies to upper or lower truncation given that $0 \leq \delta(\alpha) \leq 1$

16

Truncated Normal

Model: $y_i^* = \boldsymbol{x}_i'\beta + \varepsilon_i$

Observed Data: $y_i = y_i^* \mid y_i^* > 0$



• Truncated regression model:

$$E[y_i \mid y_i^* > 0, \boldsymbol{x}_i] = \boldsymbol{x}_i'\beta + \sigma\lambda_i$$

17

Censored and Truncated Data: Intuition

• To model the relation between the observed y and $\boldsymbol{x}$, we consider a latent variable y^* that is subject to censoring/truncation. A change in $\boldsymbol{x}$ affects y only through the effect of $\boldsymbol{x}$ on y^*.

- Model the true, latent variable: $y_i^* = f(\boldsymbol{x}_i) = \boldsymbol{x}_i'\beta + \varepsilon_i$
- Observed Variable: $y_i = h(y_i^*)$

• Q: What is this latent variable?

Considered the variable "Dividends paid last quarter"?

- For all the respondents with $y = 0$ (*left censored from below),* we think of a latent variable for "excess cash" that underlies "dividends paid last quarter." Extremely cash poor would pay negative dividends if that were possible.

Censored and Truncated Data: Problems

• In the presence of censoring/truncation, we have a dependent variable, y_i, with special attribute(s):

(1) constrained and

(2) clustered of observations at the constraint.

Examples:
- Consumption (1, not 2)
- Wage changes (2, not 1)
- Labor supply (1 & 2)

• These attributes create problems for the linear model.

Censored and Truncated Data: Problems

• Censoring in a regression framework (from Ruud).

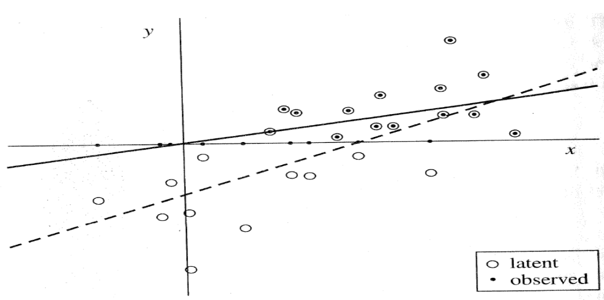


Figure 28.2 Censored regression.

• If y is constrained and if there is clustering
– OLS on the complete sample biased and inconsistent.
– OLS on the unclustered part biased and inconsistent.

Data Censoring and Corner Solutions

• The dependent variable is partly continuous, but with positive mass at one or more points.

Two different cases:

(1) **Data censoring** (from above or below)

- Surveys for wealth (upper category $250,000 or more)
- Duration in unemployment
- Demand for stadium tickets

⇒ y_i^* is not observed because of constraints, selection technique or measuring method (data problem)

- We are interested in the effect of x_k on y_i^*: E($y_i^*|\boldsymbol{x}_i$).

⇒ $y_i^* = \boldsymbol{x}_i'\beta + \varepsilon_i$, β_k measures the effect of a change of x_k on y_i^*.

Data Censoring and Corner Solutions

(2) **Corner solutions**: A significant fraction the data has zero value.

- Hours worked by married women: Many married women do not work –i.e, zero worked hours are reported.
- Luxury goods, charitable donations, alcohol consumption, etc.

⇒ y_i^*cannot be observed because of the nature of topic.

- Contrast to data censoring: Observing the dependent variable is not a problem (For modeling, we will use a latent model, y_i^*.)

- We are interested in E($y_i|\boldsymbol{x}_i$), we want to study the effect of a change in education on y, hours worked by married women.

⇒ If $y_i^* = \boldsymbol{x}_i'\beta + \varepsilon_i$, β_k is not what we are interested.

Corner Solutions

- These data sets are typical of microeconomics. We think of these data sets as a result of utility maximization problems with two decisions:

(1) Do something or not (work or not, buy IBM or not, donate or not)

⇒ The **participation** decision: A binary choice problem (y_i = 0 or y_i >0).

(2) How much we do something (how many hours married women work, how much to invest on IBM, how much to donate to charity).

⇒ The **amount** decision: This a truncated sample problem (how much y_i is, if y_i > 0).

Note: Last class, we have gone over this type of data, when the decision amount was a "count." (a hurdle model).

Tobit Model (Censored Normal Regression)

Example: We are interested on the effect of education, $\boldsymbol{x}$, on the married women's hours worked, y_i.

- A model for the latent variable y*, which is only partially observed:

$$y_i^* = \beta_0 + \beta_1\, x_i + \varepsilon_i = \boldsymbol{x}_i'\beta + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

- Data (truncated sample)
  - If $y_i^* > 0$ ⇒ y_i = Actual hours worked = $y_i^* = \boldsymbol{x}_i'\beta + \varepsilon_i$.
  - If $y_i^* \leq 0$ ⇒ $y_i = 0$ (y_i^* can be negative, but if it is, y_i=0)

- Probability Model -- $\varepsilon_i \sim N(0, \sigma^2)$
  - Prob(y=0|$\boldsymbol{x}$) = P($y^* \leq 0$|$\boldsymbol{x}$) = P[($y^* - \mathbf{X}\beta)/\sigma \leq (0 - \mathbf{X}\beta)/\sigma$|$\boldsymbol{x}$] = P($z \leq -\mathbf{X}\beta/\sigma$|$\boldsymbol{x}$) = $\Phi(-\mathbf{X}\beta/\sigma) = 1 - \Phi(\mathbf{X}\beta/\sigma)$
  - Prob(y>0|$\boldsymbol{x}$) = P($y^* > 0$|$\boldsymbol{x}$) $= 1 - \Phi(-\mathbf{X}\beta/\sigma) = \Phi(\mathbf{X}\beta/\sigma)$

24

Tobit Model (Censored Normal Regression)

- Expectations of interest:
  - Unconditional Expectation

$$E[y_i \,|\, \boldsymbol{x}_i] = P(y_i > 0 \,|\, \boldsymbol{x}_i) * E[y_i \,|\, y_i > 0, \boldsymbol{x}_i] + \text{Prob}(y_i = 0 \,|\, \boldsymbol{x}_i) * 0$$
$$= P(y_i > 0 \,|\, \boldsymbol{x}_i) * E[y_i \,|\, y_i > 0, \boldsymbol{x}_i]$$
$$= \Phi(\boldsymbol{x}_i'\boldsymbol{\beta}/\sigma) * E[y_i \,|\, y_i > 0, \boldsymbol{x}_i]$$

-Conditional Expectation (Recall: $E[y_i \,|\, y_i^* > c, \boldsymbol{x}_i] = \mu^* + \sigma\, \lambda(\alpha)$)

$$E[y_i \,|\, y_i > 0, \boldsymbol{x}_i] = \boldsymbol{x}_i'\boldsymbol{\beta} + \sigma\, \lambda(\boldsymbol{x}_i'\boldsymbol{\beta})$$

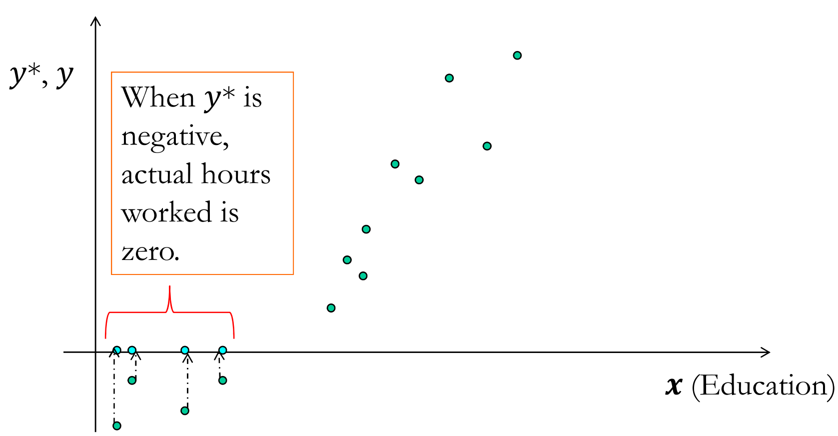
Remark: The presented Tobit model –also called *Type I Tobit Model*- can be written as a combination of two models:

(1) A Probit model: It determines whether $y = 0$ (No) or $y > 0$ (Yes).

(2) A truncated regression model for $y > 0$.

25

Tobit Model – Sample Data



- Expectations of interest:

$E[y_i \,|\, \boldsymbol{x}_i]$, $E[y_i |\, y_i > 0, \boldsymbol{x}_i]$.

$E[y_i^* | \boldsymbol{x}_i]$

26

Tobit Model - OLS

• Suppose we do OLS, only for the part of the sample with y_i > 0. The estimated model:

$$y_i = \boldsymbol{x}_i'\beta + \upsilon_i \qquad \text{for } y_i > 0$$

- Let's look at the density of υ_i, $f_\upsilon(.)$, which must integrate to 1:

$$\int_{-x_i'\beta}^{\infty} f_\upsilon(\eta)\,d\eta = 1$$

- The ε_i's density, normal by assumption in the Tobit Model:

$$\int_{-x_i'\beta}^{\infty} f_\varepsilon(\eta)\,d\eta = F_i = \int_{-\infty}^{x_i'\beta} f_\varepsilon(\eta)\,d\eta = \int_{-\infty}^{x_i'\beta} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{\eta}{\sigma})^2}$$

- Then, $f_\upsilon(.)$ can be written as:

$$f_\upsilon = F_i^{-1} f_\varepsilon = F_i^{-1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{\upsilon_i}{\sigma})^2}$$

27

Tobit Model - OLS

• The pdf of υ_i and the pdf of ε_i are different. In particular,
The ε_i's density, normal by assumption in the Tobit Model:

$$E[\upsilon_i] = \int_{-x_i'\beta}^{\infty} \eta f_\upsilon(\eta)\,d\eta = F_i^{-1}\int_{-x_i'\beta}^{\infty} \eta f_\varepsilon(\eta)\,d\eta$$

$$= F_i^{-1}\int_{-x_i'\beta}^{\infty} \eta \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{\eta}{\sigma})^2}\,d\eta = F_i^{-1}\int_{-x_i'\beta}^{\infty} \frac{\eta}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\eta}{\sigma})^2}\,d\eta$$

$$= F_i^{-1}[-\sigma\, f_\varepsilon(\eta)]_{-x_i'\beta}^{\infty} \qquad \text{integration by substitution}$$

$$= \sigma\, F_i^{-1} f_i \neq 0 \qquad (f_i = f_\varepsilon(x_i'\beta))$$

- Then, E[$\upsilon_i\,|\,\boldsymbol{x}_i$] = σ F_i^{-1} f_i = σ $\lambda(\boldsymbol{x}_i'\boldsymbol{\beta})$ ≠ 0 (& it depends on $\boldsymbol{x}_i'\boldsymbol{\beta}$)

⇒ E[$y_i\,|\,y_i > 0, \boldsymbol{x}_i'\boldsymbol{\beta}$] = $\boldsymbol{x}_i'\boldsymbol{\beta} + \sigma\,\lambda(\boldsymbol{x}_i'\boldsymbol{\beta})$

⇒ OLS in truncated part is *biased* (omitted variables problem).

28

Tobit Model - OLS

• OLS in truncated part only is *biased*. We have an omitted variables problem. With a bit more work, we can show OLS is inconsistent too (show it!)

• Also $E[\upsilon_i{}^2|\mathbf{x}] = \sigma^2 - \sigma^2\, \boldsymbol{x}_i'\boldsymbol{\beta} * \lambda(\boldsymbol{x}_i'\boldsymbol{\beta})$

⇒ The error term, υ, not only has non-zero mean but it is also heteroskedastic.

29

Tobit Model - NLLS

• Now, we can write

$$y_i = E[y_i|y_i > 0, \boldsymbol{x}_i] + \varepsilon_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \sigma\,\lambda(\boldsymbol{x}_i'\boldsymbol{\beta}) + \varepsilon_i, \quad \text{for } y_i > 0.$$

- There is a non-linear relation between $\boldsymbol{x}_i$ on y_i. NLLS is a possibility here, though we need a consistent estimator of $\boldsymbol{\beta}$ to evaluate $\lambda(\boldsymbol{x}_i'\beta)$ and it is not clear where to get it. Since we have heteroscedasticity, we need to allow for it. Weighted NLLS may work well.

<u>Note</u>: A non-linear relation appears even if all observations are used (positive and negative values of y_i):

$$E[y_i|\boldsymbol{x}_i] = P(y_i^*>0|\boldsymbol{x}_i) * E(y_i^*|y_i > 0, \boldsymbol{x}_i) = F_i * (\boldsymbol{x}_i'\beta + \sigma\, F_i^{-1} f_i)$$
$$= \Phi(\boldsymbol{x}_i'\boldsymbol{\beta}) * \{\boldsymbol{x}_i'\boldsymbol{\beta} + \sigma\,\phi(\boldsymbol{x}_i'\boldsymbol{\beta})\}$$

30

Tobit Model: Estimation

- Given our assumption for ε_i, we do ML estimation.
- For women who are working, we have $y_i^* > 0 \quad \Rightarrow y_i > 0$. Then,

$$\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i)$$

$\Rightarrow$ Likelihood function for a working woman is given by:

$$L_i = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{\sigma^2}} = \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma})^2}$$

$$= \frac{1}{\sigma}\, \phi(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma})$$

- For women who are not working, we have $y_i^* \leq 0 \quad \Rightarrow \; y_i = 0$.

$$\Rightarrow \quad L_i = P(y_i^* \leq 0) = P(\beta_0 + \beta_1 x_i + \varepsilon_i \leq 0)$$

$$= P(\varepsilon_i \leq -(\beta_0 + \beta_1 x_i)) = P(\frac{\varepsilon_i}{\sigma} \leq -\frac{\beta_0 + \beta_1 x_i}{\sigma})$$

$$= \Phi\left(-\frac{\beta_0 + \beta_1 x_i}{\sigma}\right) = 1 - \Phi\left(\frac{\beta_0 + \beta_1 x_i}{\sigma}\right)$$

31

Tobit Model: Estimation

- Summary,

$$L_i = \frac{1}{\sigma}\, \phi(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma}) \qquad \text{if } y_i^* > 0.$$

$$= 1 - \Phi\left(\frac{\beta_0 + \beta_1 x_i}{\sigma}\right) \qquad \text{if } y_i^* \leq 0.$$

- We have a combination of a pdf (for the observed part of the distribution) and a CDF (for the truncated part of the distribution): a linear part and a Probit part.

- Let D_i be a dummy variable that takes 1 if $y_i > 0$. Then, the above likelihood for consumer i can be written as:

$$L_i = \{\frac{1}{\sigma}\, \phi\left(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma}\right)\}^{D_i} * \{1 - \Phi\left(\frac{\beta_0 + \beta_1 x_i}{\sigma}\right)\}^{1-D_i}$$

32

Tobit Model: Estimation

- The likelihood function, L, for the whole sample is:

$$L(\beta_0, \beta_1, \sigma) = \prod_i \{\tfrac{1}{\sigma}\, \phi\left(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma}\right)\}^{D_i} * \{1 - \Phi\left(\frac{\beta_0 + \beta_1 x_i}{\sigma}\right)\}^{1-D_i}$$

- The values of β_0, β_1 and σ that maximize the likelihood function are the Tobit estimators of the parameters.

- As usual, we work with the Log(L):

$$L(\beta_0,\beta_1,\sigma) = \sum_{i=1}^{n} D_i \log\left[\frac{1}{\sigma}\phi(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma})\right] + (1-D_i)\log\left[1-\Phi(\frac{\beta_0+\beta_1 x_i}{\sigma})\right]$$

$$= \frac{N}{2}[\log(\sigma^2) + \log(2\pi)] +$$

$$+ \sum_{i=1}^{n} D_i\left[-\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^2} + (1-D_i)\log\left[1-\Phi(\frac{\beta_0+\beta_1 x_i}{\sigma})\right]\right]$$

33

Tobit Model: Estimation – Information Matrix

- Amemiya (1973) presents the following representation for the information matrix:

$$\underset{((K+1)\,x\,(K+1))}{I(\Theta)} = \left[\begin{array}{c|c} \underset{(K\,x\,K)}{\sum_{t=1}^{T} a_i X_i X_i'} & \underset{(K\,x\,1)}{\sum_{t=1}^{T} b_i X_i} \\ \hline \underset{(1\,x\,K)}{\sum_{t=1}^{T} b_i X_i'} & \underset{(1\,x\,1)}{\sum_{t=1}^{T} c_i} \end{array}\right]$$

where:

$$z_i = \frac{X_i'\beta}{\sigma} \qquad a_i = \frac{-1}{\sigma^2}\left(z_i f(z_i) - \frac{f(z_i)^2}{1-F(z_i)} - F(z_i)\right)$$

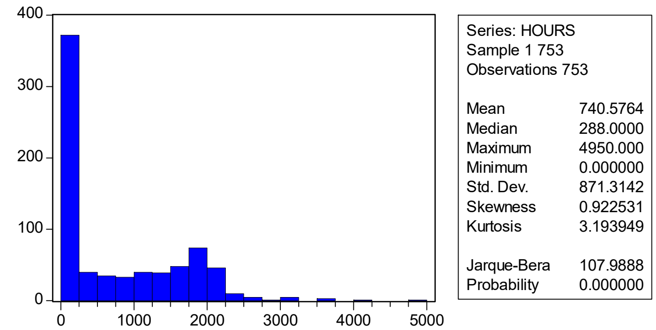
$$b_i = \frac{1}{2\sigma^3}\left(z_i^2 f(z_i) + f(z_i) - \frac{z_i f(z_i)^2}{1-F(z_i)}\right)$$

$$c_i = -\frac{1}{4\sigma^4}\left(z_i^3 f(z_i) + z_i f(z_i) - \frac{z_i^2 f(z_i)^2}{1-F(z_i)} - 2F(z_i)\right)$$

34

Tobit Model: Application I – Female Labor Supply

- **Corner Solution Case – DATA:** Female labor supply (mroz.wf1)



Variable	Obs	Mean	Std. Dev.	Min	Max
hours	753	740.5764	871.3142	0	4950
nwifeinc	753	20.12896	11.6348	-.0290575	96
educ	753	12.28685	2.280246	5	17
exper	753	10.63081	8.06913	0	45
age	753	42.53785	8.072574	30	60
kidsge6	753	1.353254	1.319874	0	8
kidslt6	753	.2377158	.523959	0	3

Tobit Model : Application I – OLS (all y)

- **OLS whole sample** (N = 753)

Dependent Variable: HOURS
Method: Least Squares
Included observations: 753
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
NWIFEINC	-3.446636	2.240662	-1.538222	0.1244
EDUC	28.76112	13.03905	2.205768	0.0277
EXPER	65.67251	10.79419	6.084062	0.0000
EXPER^2	-0.700494	0.372013	-1.882983	0.0601
AGE	-30.51163	4.244791	-7.188018	0.0000
KIDSLT6	-442.0899	57.46384	-7.693359	0.0000
KIDSGE6	-32.77923	22.80238	-1.437535	0.1510
C	1330.482	274.8776	4.840273	0.0000

Tobit Model : Application I – OLS (y>0)

- **OLS subsample** – y_i (hours worked) > 0 (N = 428)

Dependent Variable: HOURS
Method: Least Squares
Sample: 1 753 IF HOURS>0
Included observations: 428
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
NWIFEINC	0.443851	3.108704	0.142777	0.8865
EDUC	-22.78841	16.09281	-1.416061	0.1575
EXPER	47.00509	15.38725	3.054808	0.0024
EXPER^2	-0.513644	0.417384	-1.230627	0.2192
AGE	-19.66352	5.845279	-3.364001	0.0008
KIDSLT6	-305.7209	125.4802	-2.436407	0.0152
KIDSGE6	-72.36673	31.28480	-2.313159	0.0212
C	2056.643	351.4502	5.851875	0.0000

Tobit Model : Application I – Tobit

- **Tobit whole sample** (N = 753)

Dependent Variable: HOURS
Method: ML - Censored Normal (TOBIT) (Quadratic hill climbing)
Included observations: 753
Left censoring (value) at zero

	Coefficient	Std. Error	z-Statistic	Prob.
NWIFEINC	-8.814243	4.459100	-1.976687	0.0481
EDUC	80.64561	21.58324	3.736493	0.0002
EXPER	131.5643	17.27939	7.613943	0.0000
EXPER^2	-1.864158	0.537662	-3.467155	0.0005
AGE	-54.40501	7.418502	-7.333693	0.0000
KIDSLT6	-894.0217	111.8780	-7.991039	0.0000
KIDSGE6	-16.21800	38.64139	-0.419705	0.6747
C	965.3053	446.4361	2.162247	0.0306
Error Distribution				
SCALE:C(9)	1122.022	41.57910	26.98523	0.0000

Left censored obs	325	Right censored obs	0
Uncensored obs	428	Total obs	753

Partial Effects (marginal effects)

• The estimated parameters β_k measures the effect of x_k on y^*. But in corner solutions, we are interested in the effect of x_k on actual y_i.

• We calculate partial effects based on two results already derived:

- For positive y's –i.e., $y_i > 0$:

$E[y_i \mid y_i>0, \boldsymbol{x}_i'\boldsymbol{\beta}] = \boldsymbol{x}_i'\boldsymbol{\beta} + \sigma\,\lambda(\boldsymbol{x}_i'\boldsymbol{\beta})$

- For all y's –i.e., $y_i \geq 0$:

$E[y_i \mid \boldsymbol{x}_i'\boldsymbol{\beta}] = P(y_i^*>0 \mid \boldsymbol{x}_i) * E(y_i^* \mid y_i>0, \boldsymbol{x}_i) = F_i\ [\boldsymbol{x}_i'\boldsymbol{\beta} + \sigma\,\lambda(\boldsymbol{x}_i'\boldsymbol{\beta})]$

• The partial effects are given by

(1) $$\frac{\partial E(y \mid y>0, x)}{\partial x_k} = \beta_k \left\{ 1 - \lambda(\frac{x'\beta}{\sigma}) \left[\frac{x'\beta}{\sigma} + \lambda(\frac{x'\beta}{\sigma}) \right] \right\} = \beta_k (1 - \delta_k)$$

(1) measures the effect of an x_k change on y for working women.

39

Partial Effects (marginal effects)

$\Rightarrow \beta_k$ overstates the marginal impact of a change in x_k

(2) $$\frac{\delta E[y_i \mid \boldsymbol{x}_i'\boldsymbol{\beta}]}{\delta x} = \beta_k\ \phi(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma})$$ (see derivation in Greene)

(2) measures the overall effect of an x_k change on hours worked.

• Both partial effects depend on $\boldsymbol{x}$. Thus, they vary by person.

• We are interested in the overall effect rather than the effect for a specific person in the data. Two ways to do this computation:

- At the sample average: Plug the mean of $\boldsymbol{x}$ in the above formula.
- Average of partial effects: Compute the partial effect for each individual in the data. Then, compute the average.

40

Partial Effects – Application I (y>0)

```
. 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

. ****************************
. *Compute the Partial effect *
. *at average of educ         *
. *on hours for working women *
. *manually                   *
. ****************************
.   predict xbeta, xb

.   egen avxbeta=mean(xbeta)

.   gen avxbsig=avxbeta/_b[/sigma]

.   gen lambda=normalden(avxbsig)/normal(avxbsig)

.   gen partial=_b[educ]*(1-lambda*(avxbsig+lambda))

.   su partial
```

Variable	Obs	Mean	Std. Dev.	Min	Max
partial	753	34.27517	0	34.27517	34.27517

Partial effect at average for *working women*: Computing manually.

41

Partial Effects – Application I (y>0)

```
. 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
```


```
. ***********************************
. * Compute the partial effect      *
. * at average of educ on hours     *
. * for working women automatically *
. ***********************************
.   mfx, predict(e(0,.)) varlist(educ)

Marginal effects after tobit
      y  = E(hours|hours>0) (predict, e(0,.))
         =  1012.0327
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
educ	34.27517	9.11708	3.76	0.000	16.406	52.1443	12.2869

Partial effect at average for *working women*: Compute automatically.

42

Partial Effects – Application I (all y)

```
.
. ****************************************
. *Compute the Partial effect at average  *
. *of education for the entire observation*
. *manually                               *
. ****************************************
.
. gen partial_all=_b[educ]*normal(avxbsig)

. su partial_all

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 partial_all |       753    48.73409           0   48.73409   48.73409
```

Partial effect at average for *all observations*: Compute manually.

```
. ****************************************
. *Compute the Partial effect at average  *
. *of education for the entire observation*
. *automatically                          *
. ****************************************
.
. mfx, predict(ystar(0,.)) varlist(educ)

Marginal effects after tobit
      y  = E(hours*|hours>0) (predict, ystar(0,.))
         =  611.57078
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
    educ |   48.73409      12.963    3.76   0.000   23.3263  74.1419   12.2869
------------------------------------------------------------------------------
```

Partial effect at average for *all observations*: Compute automatically

43

Partial Effects – Application I (all y)

• Now, we can compare the marginal effect of education on actual hours worked.

• We compare OLS (whole sample) and Tobit estimates, on the basis of the marginal effect of education actual y_i, *for an average individual*:

OLS	**TOBIT**
$\hat{\beta}_{k,OLS}$	$\underbrace{\underset{80.65}{\beta_k}\ \underset{0.604}{\underbrace{\Phi\left(\frac{\bar{x}'\beta}{\sigma}\right)}}}_{48.73}$
28.76	48.73

→ OLS <u>underestimates</u> the effect of education on the labor supply (in the average of the explanatory variables).

<u>Interpretation</u>: On average, an additional year of education increases the labor supply by 48.7 hours (for an average individual).

Tobit Model: Heteroscedasticity

- In a regression model, we scale the observations by their standard deviation (x_i/σ_i) transforming the model back to CLM framework

- In the Tobit model, we naturally work with the likelihood. The Log L for the homoscedastic Tobit model is:

$$L(\beta_0,\beta_1,\sigma) = \frac{N}{2}[\log(\sigma^2) + \log(2\pi)] +$$
$$+\sum_{i=1}^{n}\left[-D_i\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^2} + (1-D_i)\log\left[1-\Phi(\frac{\beta_0+\beta_1 x_i}{\sigma})\right]\right]$$

- Introducing heteroscedasticity in the Log L:

$$L(\beta_0,\beta_1,\sigma_1,\sigma_2,...,\sigma_N) = \frac{N}{2}\log(2\pi) +$$
$$+\sum_{i=1}^{n}\left[D_i[\log(\sigma_i) - \frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma_i^2}] + (1-D_i)\log[1-\Phi(\frac{\beta_0+\beta_1 x_i}{\sigma_i})]\right]$$

45

Tobit Model: Heteroscedasticity

- Now, we went from k+1 parameters to k+N parameters. Impossible to estimate with N observations.

- Usual solution: Model heteroscedasticity, dependent on a few parameters: $\sigma_i^2 = \sigma_i(\alpha)$.

Example: Exponential: $\sigma_i^2 = \exp(\mathbf{z}_i'\alpha)$. Then,

$$L(\beta_0,\beta_1,\alpha) = \frac{N}{2}\log(2\pi) +$$
$$+\sum_{i=1}^{n}\left[D_i[z_i'\alpha - \frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\exp(z_i'\alpha)}] + (1-D_i)\log[1-\Phi(\frac{\beta_0+\beta_1 x_i}{\sqrt{\exp(z_i'\alpha)}})]\right]$$

- The marginal effects get more complicated under the heteroscedastic Tobit model: an exogenous variable, say income, could impact both numerator and denominator of the standardization ratio, $X_i\beta/\sigma_i$

46

Heteroscedasticity – Partial Effects

• Partial effects get more complicated under the heteroscedastic Tobit model: an exogenous variable, say income, could impact both numerator, $\boldsymbol{x}_i'\beta$, and denominator σ_i. Ambiguous signs are possible.

• Suppose we have $\boldsymbol{w}_j$ affecting both $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$. Then,

$$r_i = \frac{x_i'\beta_i}{\sigma_i} = \frac{x_i'\beta_i}{\exp(z_i'\alpha/2)}$$

$$\frac{\partial r_i}{\partial w_{ji}} = \frac{\beta_i\sigma_i - (x_i'\beta_i)(\alpha_i/2)\sigma_i}{\sigma_i^2} = \frac{\beta_i - (x_i'\beta_i)(\alpha_i/2)}{\sigma_i}$$

$$\frac{\partial \Phi(r_i)}{\partial w_{ji}} = \phi(r_i)\frac{\partial r_i}{\partial w_{ji}} = \phi(r_i)\frac{\beta_i - (x_i'\beta_i)(\alpha_i/2)}{\sigma_i}$$

$$\frac{\partial E[y_i \mid y_i > 0]}{\partial w_{ji}} = \beta_i - \sigma_i \frac{\partial r_i}{\partial w_{ji}}\left(r_i \frac{\phi(r_i)}{\Phi(r_i)} + \frac{\phi(r_i)^2}{\Phi(r_i)^2} \right)$$

Heteroscedasticity – Partial Effects - Application

• Canadian FAFH expenditures: 9,767 HH's, 21.2% with \$0 expenditures.

• Dependent variable is bi-weekly FAFH expenditures
Exogenous Variables: HHInc, Kids Present?, FullTime? Provincial Dummy Variables.

• $\sigma_i^2 = \exp(\gamma_0 + \gamma_1 \text{Income}_i + \gamma_2 \text{Fulltime}_i + \gamma_3 \text{Quebec}_i)$

Elasticity	Homo.		Hetero.	
	Value	S.E.	Value	S.E
Φ(z)	0.259	0.008	0.210	0.011
E(y|y>0)	0.284	0.009	0.395	0.010
E(y)	0.544	0.017	0.606	0.020

Tobit Model – Type II

• Different ways of thinking about how the latent variable and the observed variable interact produce different Tobit Models.

• The Type I Tobit Model presents a simple relation:

- $y_i = 0$ if $y_i^* = \boldsymbol{x}_i'\beta + \varepsilon_i \leq 0$
  $= y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$ if $y_i^* = \boldsymbol{x}_i'\beta + \varepsilon_i > 0$

The effect of the $\boldsymbol{x}$'s on the probability that an observation is censored and the effect on the conditional mean of the non-censored observations are the same: $\boldsymbol{\beta}$.

• The Type II Tobit Model presents a more complex relation:

- $y_i = 0$ if $y_i^* = \boldsymbol{x}_i'\alpha + \varepsilon_{1,i} \leq 0$, $\varepsilon_{1,i} \sim N(0, 1)$
  $= \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_{2,i}$ if $y_i^* = \boldsymbol{x}_i'\alpha + \varepsilon_{2,i} > 0$, $\varepsilon_{2,i} \sim N(0, \sigma_2^2)$

Now, we have different effects of the $\boldsymbol{x}$'s.

49

Tobit Model – Type II

• The Type II Tobit Model:

- $y_i = 0$ if $y_i^* = \boldsymbol{x}_i'\alpha + \varepsilon_{1,i} \leq 0$, $\varepsilon_{1,i} \sim N(0, 1)$
  $= \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_{2,i}$ if $y_i^* = \boldsymbol{x}_i'\alpha + \varepsilon_{2,i} > 0$, $\varepsilon_{2,i} \sim N(0, \sigma_2^2)$

- A more flexible model. **X** can have an effect on the decision to participate (Probit part) and a different effect on the amount decision (truncated regression).

- Type I is a special case: $\varepsilon_{2,i} = \varepsilon_{1,i}$ and $\alpha = \boldsymbol{\beta}$.

Example: Age affects the decision to donate to charity. But it can have a different effect on the amount donated. We may find that age has a positive effect on the decision to donate, but given a positive donation, younger individuals donate more than older individuals.

50

Tobit Model – Type II

• The Tobit Model assumes a bivariate normal distribution for ($\varepsilon_{1,i}$;$\varepsilon_{2,i}$); with covariance given by σ_{12}(= $\rho\ \sigma_1 \sigma_2$.).

- Conditional expectation:

$$E[y_i \,|\, y_i > 0, \boldsymbol{x}_i] = \boldsymbol{x}_i{}'\boldsymbol{\beta} + \sigma_{12}\ \lambda(\boldsymbol{x}_i{}'\boldsymbol{\alpha})$$

- Unconditional Expectation

$$\begin{aligned} E[y_i \,|\, \boldsymbol{x}_i] &= \text{Prob}(y_i{>}0 \,|\, \boldsymbol{x}_i) * E[y_i \,|\, y_i{>}0, \boldsymbol{x}_i] + \text{Prob}(y_i{>}0 \,|\, \boldsymbol{x}_i) * 0 \\ &= \text{Prob}(y_i > 0 \,|\, \boldsymbol{x}_i) * E[y_i \,|\, y_i > 0, \boldsymbol{x}_i] \\ &= \Phi(\boldsymbol{x}_i{}'\boldsymbol{\alpha}) * [\boldsymbol{x}_i{}'\boldsymbol{\beta} + \sigma_{12}\ \lambda(\boldsymbol{x}_i{}'\boldsymbol{\alpha})] \end{aligned}$$

Note: This model is known as the Heckman selection model, or the Type II Tobit model (Amemiya), or the probit selection model (Wooldridge).

51