Corner solutions and censored regression models

- Censored regression models are used in two economic different contexts:
  - data censoring
  - corner solutions outcomes

- Data censoring: part of the sample is coded as one value. Usually we have actual values of the variable up to the trashold and for the rest of the sample we only know that it is above the trashold.

Example 1. *(Wooldrige). We want to analyze the determinants of wealth. The wealth variable is however constructed in such a way that we have the actual wealth (wealth*) of the family is below \$200,000 and \$200,000 if actual wealth is above \$200,000. Assume that actual wealth fallows classical* *regression model*

$$\textit{wealth}^* = \boldsymbol{x}\boldsymbol{\beta} + u, \quad u|\,\boldsymbol{x} \sim Normal\left(0, \boldsymbol{\sigma}^2\right)$$

But we only have information on

$$\textit{wealth} = \min\left(200, \textit{wealth}^*\right) = \min\left(200, \boldsymbol{x}\boldsymbol{\beta} + u\right)$$

We have to estimate the parameter $\boldsymbol{\beta}$ on the basis of the variable wealth and not the latent variable wealth*

- For many economic problems the outcomes of the agent decision is either zero or a positive number (e.g. amount spend for a durable good, insurance coverage chosen, firm expenditures for R&D etc.)

- So we will observe some positive values of a variable but also a number of zeros for these agents for whom the optimal choice was the corner solution $y = 0$.

- This kind of response variable is called corner solutions outcomes.

- In this case the classical regression model is usually inappropriate because

  - in linear regression $y = \boldsymbol{x\beta} + \boldsymbol{\varepsilon}$ there is no way to guarantee that all predicted values are positive
  - in nonlinear regression $y = \exp(\boldsymbol{x\beta}) + \boldsymbol{\varepsilon}$, we will usually have heteroskedasticity
  - in neither of these regression it is possible to obtain two interesting features of the distribution of y:
    * $P(y = 0|\, \boldsymbol{x})$ - probability of corner solution conditional on $\boldsymbol{x}$
    * $\mathrm{E}(y|\, \boldsymbol{x}, y > 0)$ - conditional expectation of y given that y is positive (is not corner solution)

- It is important to notice that although the data sets resulting from data censoring and corner solutions outcomes can look very similar the underlying cause is conceptually very different - for corner solutions the issue is not observability (in this case we always observe actual decisions)

- The statistical model used in this context is so called *standard censored Tobit model* or *type I Tobit model*. It is defined in the following way:

$$\begin{aligned} y_i^* &= \boldsymbol{x}_i\boldsymbol{\beta} + u_i \qquad\qquad u_i|\,\boldsymbol{x}_i \sim Normal\left(0, \boldsymbol{\sigma}^2\right) \\ y_i &= \max\left(0, y_i^*\right) \end{aligned}$$

Example 2. *(Wooldridge) wealth cont. This problem can be easily transformed to the standard model as*

$$-\left(\textit{wealth} - 200\right) = \max\left(0, -200 - \boldsymbol{x\beta} - u\right)$$

- Different features of the model are of interest depending on application:

- for true data censoring we are usually interested in $\mathrm{E}\left(y^{*}\middle|\, \boldsymbol{x}\right) = \boldsymbol{x\beta}$ (model determining the latent variable)
- for corner response variable we are interested in $P\left(y = 0\middle|\, \boldsymbol{x}\right)$ or $\mathrm{E}\left(y\middle|\, \boldsymbol{x}, y > 0\right)$ as latent variable usually has no sensible interpretation in this context.

- The variable y^{*} in standard regression model should be roughly continuous and have homoscedastic normal distribution (for data censoring cases the logaritmic transformation can be used to make this assumption more plausible)

- The expected value of y is derived in the following way:

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

where $\boldsymbol{\lambda}\left(c\right)=\frac{\phi(c)}{\Phi(c)}$ is called *inverse Mills ratio*.

- If x_j is continuous than the partial effect of x_j on $\mathrm{E}(y|\,\boldsymbol{x},y>0)$ is given by

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

  as it si possible to prove that $1-\boldsymbol{\lambda}(\boldsymbol{x\beta}/\,\boldsymbol{\sigma})\left[\boldsymbol{x\beta}/\,\boldsymbol{\sigma}+\boldsymbol{\lambda}(\boldsymbol{x\beta}/\,\boldsymbol{\sigma})\right]>0$ this partial effect has the same sign as $\boldsymbol{\beta}_j$.

- The partial effect of x_j on $P(y>0|\,\boldsymbol{x})$ is equal to

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

- The partial effect of x_j on $\mathrm{E}(y|\,\boldsymbol{x})$ is given by

$$\begin{aligned}\frac{\partial\,\mathrm{E}(y|\,\boldsymbol{x})}{\partial x_j} &= \frac{\partial P(y>0|\,\boldsymbol{x})}{\partial x_j}\mathrm{E}(y|\,\boldsymbol{x},y>0)+P(y>0|\,\boldsymbol{x})\frac{\partial\,\mathrm{E}(y|\,\boldsymbol{x},y>0)}{\partial x_j}\\ &= \Phi(\boldsymbol{x\beta}/\,\boldsymbol{\sigma})\,\boldsymbol{\beta}_j\end{aligned}$$

- The interpretation of this result is the following: the $\Phi(\boldsymbol{x\beta}/\,\boldsymbol{\sigma})$ is the scale factor equal to probability of observing positive outcome. If this probability is close to one than the effect of censoring/corner solutions is small and practically the same as linear model (usually we only consider $\Phi(\overline{\boldsymbol{x}}\boldsymbol{\beta}/\,\boldsymbol{\sigma})$). In most interesting tobit applications $\Phi(\overline{\boldsymbol{x}}\boldsymbol{\beta}/\,\boldsymbol{\sigma})$ is significantly smaller than unity, and number of zeros in the sample substantial.

- The relative partial effects of x_j and x_h on $\mathrm{E}(y|\,\boldsymbol{x},y>0)$, $\mathrm{E}(y|\,\boldsymbol{x})$ and $P(y>0|\,\boldsymbol{x})$ are all equal to $\boldsymbol{\beta}_j/\,\boldsymbol{\beta}_h$

- The OLS estimator of $\boldsymbol{\beta}$ calculated for $y > 0$ is inconsistent as

$$y_i = \boldsymbol{x\beta} + \boldsymbol{\sigma\lambda}\left(\boldsymbol{x\beta}/\,\boldsymbol{\sigma}\right) + e_i$$
$$\mathrm{E}\left(e_i|\,\boldsymbol{x,}y > 0\right) = 0$$

and OLS is inconsistent because of the omitted variable $\boldsymbol{\lambda}\left(\boldsymbol{x\beta}/\,\boldsymbol{\sigma}\right)$ which is usually strongly correlated with $\boldsymbol{x\beta}$.

- So using OLS using uncensored observations in will usually give the inconsistent estimator of $\boldsymbol{\beta}$

- The OLS estimator of $\boldsymbol{\beta}$ calculated for all y is also inconsistent as

$$\mathrm{E}\left(y|\,\boldsymbol{x}\right) = \Phi\left(\boldsymbol{x\beta}/\,\boldsymbol{\sigma}\right)\boldsymbol{x\beta}+\boldsymbol{\sigma\phi}\left(\boldsymbol{x\beta}/\,\boldsymbol{\sigma}\right)$$

is not a linear model.

- Estimation of the tobit model. This model can be estimated with Maximum Likelihood. The density function of y is

$$
\begin{aligned}
f\left(y_i \middle| \boldsymbol{x}_i\right) &= \begin{cases} 1-\Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}/\,\boldsymbol{\sigma}\right) & \text{if} \quad y_i = 0 \\ \frac{1}{\boldsymbol{\sigma}}\boldsymbol{\phi}\left(\frac{y_i-\boldsymbol{x}_i\boldsymbol{\beta}}{\boldsymbol{\sigma}}\right) & \text{if} \quad y_i > 0 \end{cases} \\
&= \left\{1-\Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}/\,\boldsymbol{\sigma}\right)\right\}^{1[y_i=1]}\left[\frac{1}{\boldsymbol{\sigma}}\boldsymbol{\phi}\left(\frac{y_i-\boldsymbol{x}_i\boldsymbol{\beta}}{\boldsymbol{\sigma}}\right)\right]^{1[y_i>0]}
\end{aligned}
$$

Reporting of the results

- The correct way depends on the problem.
  - For true data censoring the quantities of interest are $\widehat{\boldsymbol{\beta}}_j$ and their standard errors.
  - For corner solutions applications estimated partial effects $\mathrm{E}\left(y \middle| \boldsymbol{x}\right)$, $\mathrm{E}\left(y \middle| \boldsymbol{x}, y > 0\right)$.

- Partial effects have to be calculated for some $\boldsymbol{x}$ (usually $\overline{\boldsymbol{x}}$). Standard errors of partial effects can be found using delta method.

Example 3. *(Wooldridge) Annual Hours Equation for Married Women. We estimate the reduced form of this equation as wage offer is not included (as it is probably not exogenous - depends on hours). Out of* 753 *women in the sample* 325 *work zero hours.*

Independent variable	Linear(OLS)	Tobit(MLE)
nwifeinc	-3.45 (2.54)	-8.81 (4.46)
educ	28.76 (12.95)	80.65 (21.58)
exper	65.67 (9.96)	131.56 (17.28)
exper^2	$-.700$ $(.325)$	-1.86 (0.54)
age	-30.51 (4.36)	-54.41 (7.42)
kidslt6	-442.09 (58.85)	-894.02 (111.88)
kidsge6	-32.78 (23.18)	-16.22 (38.64)
constant	$1,330.48$ (270.78)	965.31 (446.44)
Log-likelihood value		$-3,819.09$
R-squared	$.266$	$.275$
$\widehat{\boldsymbol{\sigma}}$	750.18	$1,122.02$

In order to obtain partial effects $\frac{\partial \mathrm{E}(y|x,y>0)}{\partial x_j}$ for tobit model we should multiply the estimates of $\widehat{\boldsymbol{\beta}}_j$ by factor $_j \left\{1 - \boldsymbol{\lambda}\left(\boldsymbol{x\beta}/\,\boldsymbol{\sigma}\right)\left[\boldsymbol{x\beta}/\,\boldsymbol{\sigma} + \boldsymbol{\lambda}\left(\boldsymbol{x\beta}/\,\boldsymbol{\sigma}\right)\right]\right\}$. For this problem it is equal to $.451$. For example the estimated effect of the one additional year of education on hours worked is equal to $.451 \times 80.65 = 36.4$ and of one more young child $.451 \times -894.02 = 403.2$ hours. To obtain the partial effects $\frac{\partial \mathrm{E}(y|x)}{\partial x_j}$ for tobit model we should multiply the estimates of $\widehat{\boldsymbol{\beta}}_j$ by factor $\Phi\left(\overline{\boldsymbol{x}}\boldsymbol{\beta}/\,\boldsymbol{\sigma}\right)$ equal for this problem $.645$. In most case the estimated effect is well above the OLS estimate e.g. effect of education is $.645 \times 80.65 = 52.02$ which is nearly twice bigger that 28.75.

- For tobit model R^2 was calculated as the square of correlation coefficient between y and $\widehat{y}$ (in the same way the R^2 in OLS can be calculated).

Specification issues

- **Neglected heterogeneity**

$$y = \max\left(0, \boldsymbol{x\beta} + \boldsymbol{\gamma} q + u\right), \qquad u|\,\boldsymbol{x}, q \sim Normal\left(0, \boldsymbol{\sigma}^2\right)$$

  q is not observable (omitted variable not correlated with the error term)

- If q has a normal this is not a problem, $\boldsymbol{\beta}$ and partial effects estimated correctly

- **Endogenous explanatory variable**

$$\begin{aligned} y_1 &= \max\left(0, \boldsymbol{z}_1 \boldsymbol{\beta} + \boldsymbol{\alpha}_1 y_2 + u_1\right) \\ y_2 &= \boldsymbol{z\delta}_2 + v_2 = \boldsymbol{z}_1 \boldsymbol{\delta}_{21} + \boldsymbol{z}_2 \boldsymbol{\delta}_{22} + v_2 \end{aligned}$$

The error terms u_1 and v_2 correlated and so y_2 correlated with u_1 (endogenous)

- If u_1 and v_1 bivariate normal, then $u_1 = \boldsymbol{\theta}_1 v_2 + e_1$ and $\boldsymbol{\theta}_1 = \boldsymbol{\eta}_1 / \boldsymbol{\tau}_2^2$, $\boldsymbol{\eta}_1 = \mathrm{Cov}\,(u_1, v_2)$, $\boldsymbol{\tau}_2^2 = \mathrm{Var}\,(v_2)$ and e_1 is independent of v_2.

- The the first equation can be written as

$$y_1 = \max\,(0, \boldsymbol{z}_1\,\boldsymbol{\beta} + \boldsymbol{\alpha}_1 y_2 + \boldsymbol{\theta}_1 v_2 + e_1)$$

where all the explanatory variables $\boldsymbol{z}_1$, y_2, v_2 are uncorrelated with e_1!

- But: v_2 can not be observed!

- However it can estimated as residuals $\widehat{v}_2$ from regression $y_2 = \boldsymbol{z}\boldsymbol{\delta}_2 + v_2$.

- Two step procedure

1. Estimate regression of y_2 on $\boldsymbol{z}$ obtain residuals $\widehat{v}_2$
2. Estimate tobit regression of y_1 on $\boldsymbol{z}_1$, y_2, $\widehat{v}_2$

- This procedure gives consistent estimates of $\boldsymbol{\beta}$,$\boldsymbol{\alpha}_1$ and $\boldsymbol{\theta}_1$, but in order to achieve identification we should have *at least* one variable which is in $\boldsymbol{z}$ but not in $\boldsymbol{z}_1$

- In order to test exogeneity has a very simple parametric form H_0:$\boldsymbol{\theta}_1 = 0$ - can be tested with standard t-statistics

- If $\boldsymbol{\theta}_1 \neq 0$ we can not use the standard variance matrix from the second stage regression - adjustments are needed to take into account that $\widehat{v}_2$ was estimated

Example 4. *(Wooldridge). Testing exogeneity of eduction in the hours equation. We assume that motheduc, fatheduc, huseduc influence education but not wages. First we estimate equation for education. The we add the* *residuals from this regression to the model for hours. t-statistics for the coefficient estimated for these residuals is equal to $.91$. Then we accept the hypothesis that $\boldsymbol{\theta}_1 = 0$, which implies that education is exogenous to hours.*

- **Heteroscedasticity and nonnormality**

- Generally speaking heteroscedasticity or nonnormality of the error term leads to inconsistency of tobit estimation

- There are specification tests for normality and heteroscedasticity

- If heteroscedasticity is present we normally generalize the model allowing $\boldsymbol{\sigma}_i^2 = f(\boldsymbol{x}_i\boldsymbol{\gamma})$ and estimate the generalized model with ML

- For nonnormal errors we can us LAD estimator rather than tobit

- Assume that
$$y^* = \boldsymbol{x\beta} + u, \qquad \mathrm{Med}\left(u|\,\boldsymbol{x}\right) = 0$$
as for any nondecresing function $g(\cdot)$, $\mathrm{Med}\left[g(y)\right] = g\left[\mathrm{Med}\left(y\right)\right]$ (but not so for expected value)
$$\mathrm{Med}\left(u|\,\boldsymbol{x}\right) = \mathrm{Med}\left[y - \max\left(0, \boldsymbol{x\beta}\right)|\,x\right]$$
This suggests estimating $\boldsymbol{\beta}$ by maximizing
$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left|y_i - \max\left(0, \boldsymbol{x}_i\boldsymbol{\beta}\right)\right|$$
as the unconditional estimator of median $\boldsymbol{\mu}$ is found by $\min_{\boldsymbol{\mu}} \sum_{i=1}^{n} |u_i - \boldsymbol{\mu}|$.
- Consistency follows from $M-$estimation results. Proving $\sqrt{n}$ asymptotic normality is difficult.

- **Tobit estimation in panels**

- Parameter β can be consistently estimated with pooled regression if u_{it} is independent of $\boldsymbol{x}_{it}$. However variance matrix has to be adjusted. Define unit as clusters.

- Sometimes the random effects tobit model is defined

$$
\begin{aligned}
y_{it} &= \max\left(0, \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}\right) \\
u_{it}|\, x_i, c_i &\sim Normal\left(0, \boldsymbol{\sigma}_u^2\right) \\
c_i|\, \boldsymbol{x}_i &\sim Normal\left(\boldsymbol{\psi} + \overline{\boldsymbol{x}}_i\boldsymbol{\xi}, \boldsymbol{\sigma}_a^2\right)
\end{aligned}
$$

- This model can be consistently estimated with ML.