Corner Solution Problems

(GB: Chapter 17)

Isabel Casas

Summary of models so far

- Multivariate linear regression model: $-\infty < \mathbf{y} < \infty$, continuous
  - $price \sim \beta_0 + \beta_1\ housesq + \beta_2\ plotsq + \beta_3\ room + \boldsymbol{\epsilon}$
  - $log(wage) \sim \beta_0 + \beta_1\ educ + \beta_2\ exper + \beta_3\ exper^2 + \boldsymbol{\epsilon}$
  - OLS, IV, 2SLS
- Binary decision models: $\mathbf{y} = 0$ or $\mathbf{y} = 1$, discrete
  - Latent variable $\mathbf{y}^* > 0$ or $\mathbf{y}^* \leq 0$, continuous
  - LPM, probit, logit
- Multinomial decision models: $\mathbf{y} = 0, 1, 2, \ldots, J$, discrete
  - Comparison of utility functions $\mathbf{y}^*$, continuous
  - MNL, CLM, Mixed Logit, Nested Logit.
  - Ordered probit and logit
- Censored regression models: $\mathbf{y} = 0$ or $\mathbf{y} > 0$, continuous but discrete at $\mathbf{y} = 0$

Censored Regression Models

- Two types of censored models
  - Data censoring
  - Corner solution outcomes
- Tobit model
- Expected values and marginal effects
- Estimation (MLE)
- Reporting results
- Specification issues
  - Heteroskedasticity
  - Endogeneity
- Alternatives to the Tobit model

Types of censored regression

1. **Data censoring**, e.g. **right censoring** or top coding

- family wealth: $\mathbf{y}^* = wealth^* = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
  - Only values until $200,000 are recorded.
  - Over that, only a $wealth^* = 200$ is recorded
  - $\mathbf{y} = wealth = \min(wealth^*, 200)$
- We want to estimate $E(wealth^*|\mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$
- Do you see any problem on that?
- We are actually estimating $E(wealth|\mathbf{X}) \neq \mathbf{X}\boldsymbol{\beta}$

Of course you could have a **left censoring**, or a **two sided** coding.

Two types of censored regression

1. **Data censoring**
   - When the true variable is continuous, but some values are scaled when recorded
   - Here the censoring is a feature imposed in the collection process
   - Common in many surveys
   - But less so in register data.

Q: What is the difference between censored data and truncated data?

Two types of censored regression

1. **Data censoring**
   - When the true variable is continuous, but some values are scaled when recorded
   - Here the censoring is a feature imposed in the collection process
   - Common in many surveys
   - But less so in register data.

Q: What is the difference between censored data and truncated data?

- Censored data: all values of $\mathbf{X}$ are recorded but information on $\mathbf{y}^*$ is scaled (Censored mean is different from the mean of $\mathbf{y}^*$)
- Truncated data: some values of $\mathbf{y}^*, \mathbf{X}$ are lost, we only record values over/under a threshold

Two type of censored regression

2. **Corner solution outcomes**

- $\mathbf{y}$ takes values
  - $\mathbf{y} = 0$ with positive probability and
  - $\mathbf{y} \in (0, \infty)$ is a continuous random variable
- Labour supply (hours worked)
- Wage income (many people do not have wage income)
- Firm expenditures on R&D
- Here, the censoring is a feature of the underlying behaviour.

Example corner solution outcomes

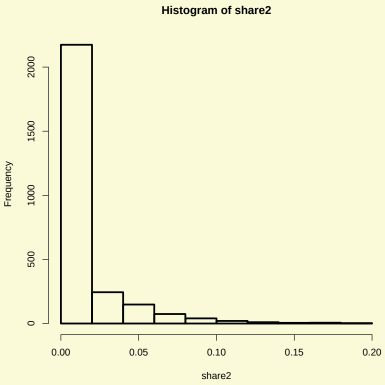
Example: tobacco.csv

bluecol:	dummy, 1 if head is blue collar worker (1)
whitecol:	dummy, 1 if head is white collar worker (1)
flanders:	dummy, 1 if living in Flanders (2)
walloon:	dummy, 1 if living in Wallonie (2)
nkids:	number of children $>$ 2 years old
nkids2:	number of children $\leq$ 2 years old
nadults:	number of adults in household
lnx:	log total expenditures
share2:	budgetshare tobacco
share1:	budgetshare alcohol
nadlnx	nadults $\times$ lnx (inter)
agelnx:	age $\times$ lnx (inter)
age:	age in brackets (0-4)
d1:	dummy, 1 if share1$>$0
d2:	dummy, 1 if share2$>$0
w1:	budgetshare alcohol ,if $>$0, missing otherwise
w2:	budgetshare tobacco ,if $>$0, missing otherwise
lnx2:	lnx squared
age2:	age squared

Example corner solution outcomes

Variable share2 (1688 observations share2=0)

```
  Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
0.00000 0.00000 0.00000 0.01224 0.01381 0.19280
```



Example corner solution outcomes

The previous example:

- The demand for tobacco is a continuous variable for values above 0, but has strictly positive probability mass at 0.
- Modelling this as a linear model is inappropriate
  - The $E(\mathbf{y}|\mathbf{X})$ is not linear for $\mathbf{y} > 0$
  - Constant partial effects are not appropriate
  - The predicted values can be negative
- We could transform it to a discrete choice model (e.g. probit)
  - y=0 if no consumption
  - y=1 if positive consumption
- But we can do better than this: Censored regression models
- A mix between a linear model and a discrete response model.

OLS?

Although the latent models are linear, the OLS regression result in inconsistent parameter estimates because the sample does not represent the population

Type I Tobit model

Both models can be given a latent variable formulation:

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon}|\mathbf{X} \sim N(0, \sigma^2)$$

$$\mathbf{y} = \max(0, \mathbf{y}^*) \qquad \Rightarrow \qquad \mathbf{y} = \begin{cases} \mathbf{y}^* & \text{if } \mathbf{y}^* > 0 \\ 0 & \text{if } \mathbf{y}^* \leq 0 \end{cases}$$

- $\mathbf{y}^*$ is normal and homokedastic and $\mathbf{y}$ is left censored

Q: How do we rewrite the wealth example (right censored) into this model? (3 minutes)

Type I Tobit model

Both models can be given a latent variable formulation:

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon}|\mathbf{X} \sim N(0, \sigma^2)$$

$$\mathbf{y} = \max(0, \mathbf{y}^*) \qquad \Rightarrow \qquad \mathbf{y} = \begin{cases} \mathbf{y}^* & \text{if } \mathbf{y}^* > 0 \\ 0 & \text{if } \mathbf{y}^* \leq 0 \end{cases}$$

- $\mathbf{y}^*$ is normal and homokedastic and $\mathbf{y}$ is left censored

Q: How do we rewrite the wealth example (right censored) into this model? (3 minutes)

$$\mathbf{y}^* = wealth = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$(wealth - 200) \leq 0 \Rightarrow -(wealth - 200) = 200 - wealth \geq 0$$

$$\mathbf{y} = \max(0, 200 - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\epsilon}) = \max(0, 200 - wealth)$$

Type I Tobit model

Both models can be given a latent variable formulation:

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon}|\mathbf{X} \sim N(0, \sigma^2)$$

$$\mathbf{y} = \max(0, \mathbf{y}^*) \qquad \Rightarrow \qquad \mathbf{y} = \begin{cases} \mathbf{y}^* & \text{if } \mathbf{y}^* > 0 \\ 0 & \text{if } \mathbf{y}^* \leq 0 \end{cases}$$

- $\mathbf{y}^*$ is normal and homokedastic and $\mathbf{y}$ is left censored
- We are interested in estimating $E(\mathbf{y}^*|\mathbf{X})$.
- However we observe $\mathbf{y}$ not $\mathbf{y}^*$!!!
- How do we interpret $\boldsymbol{\beta}$ in regard to $\mathbf{y}^*$?

Tobit model

- In data censoring models, we are interested in β in the same way than in the OLS
  - $\frac{\partial E(\mathbf{y}^*|\mathbf{X})}{\partial \mathbf{x}_k} = \frac{\partial \mathbf{X}\boldsymbol{\beta}}{\partial \mathbf{x}_k}$
- In corner solution models, we are typically interested in:
  - $E(\mathbf{y}|\mathbf{X})$
  - $E(\mathbf{y}|\mathbf{X}, \mathbf{y} > 0)$
  - β is involved in these equations
- The marginal effects of $\mathbf{X}$ on the expected value
  - For data censoring: β
  - For corner solutions models: A pondered β

Expected values

- What is the expected tobacco consumption given that the household smokes?
- I.e. the expected value of $\mathbf{y}$ given that $\mathbf{y} > 0$. This can be shown to be given by:

$$E(\mathbf{y}|\mathbf{X}, \mathbf{y} > 0) = \mathbf{X}\boldsymbol{\beta} + \sigma \frac{\phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)}{\Phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)} = \mathbf{X}\boldsymbol{\beta} + \sigma\lambda\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)$$

- It is the mean of a truncated normal distribution (truncated from the bottom) (.pdf)
- $\lambda\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right) > 0$ is called the *inverse Mill's ratio*

Truncated normal (5 minutes)

Check it out on the web:

- How is the density function of a truncated normal distribution?
- How does it look?
- How is its expected value?

Expected values

$$E(\mathbf{y}|\mathbf{X}, \mathbf{y} > 0) = \mathbf{X}\boldsymbol{\beta} + \sigma\lambda\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)$$

- This expected value is positive (as we condition on $\mathbf{y} > 0$)
- This expected value is greater than $\mathbf{X}\beta$ (as the Mill's ratio is positive)
  - For the given sample, $(\mathbf{x}_j, y_j)$, we include only observations where $\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} > 0$,
  - For a given value of $\mathbf{X}\beta$ we exclude the observations with $\boldsymbol{\epsilon}$ values such that $\Rightarrow E(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}|\mathbf{X}, \mathbf{y} > 0) > \mathbf{X}\boldsymbol{\beta}$
  - We oversample observations with positive $\boldsymbol{\epsilon}$

Expected values

What is the overall expected tobacco consumption?

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

Where:

$$\begin{aligned}
P(\mathbf{y} > 0|\mathbf{X}) =& P(\mathbf{y}^*|\mathbf{X}) = P(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} > 0|\mathbf{X}) \\
=& P(\boldsymbol{\epsilon} > -\mathbf{X}\boldsymbol{\beta}|\mathbf{X}) = P\left(\frac{\boldsymbol{\epsilon}}{\sigma} > \frac{-\mathbf{X}\boldsymbol{\beta}}{\sigma}\right) \\
=& P\left(\frac{\boldsymbol{\epsilon}}{\sigma} < \frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right) = \Phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)
\end{aligned}$$

Expected values

$$
\begin{aligned}
E(\mathbf{y}|\mathbf{X}) =& \Phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right) E(\mathbf{y}|\mathbf{X}, \mathbf{y} > 0) \\
=& \Phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right) \left[\mathbf{X}\boldsymbol{\beta} + \sigma \frac{\phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)}{\Phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)}\right] \\
=& \Phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right) \mathbf{X}\boldsymbol{\beta} + \sigma\phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)
\end{aligned}
$$

Expected values

In summary, for the corner solution outcome problem, we are interested in estimating:

$$E(\mathbf{y}|\mathbf{X}, \mathbf{y} > 0) = \mathbf{X}\boldsymbol{\beta} + \sigma\lambda\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)$$

$$E(\mathbf{y}|\mathbf{X}) = \Phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)\mathbf{X}\boldsymbol{\beta} + \sigma\phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)$$

Exercise (5 minutes)

Consider the following dependent variables:

- Hours worked during a week
- Preparation time for an exam
- Household wealth (from a survey)

1. Which of these should be modelled using a data censoring model and which should be modelled using a corner solution model?
2. Why β is not in itself interesting in a corner solution model?
3. Why are we more interested in the effects on $E(\mathbf{y}|\mathbf{X})$ and $E(\mathbf{y}|\mathbf{X}, \mathbf{y} > 0)$
4. And what can we say about the size of these expected values?

Marginal effects

For a continuous explanatory variable, it can be shown that:

$$\frac{\partial E(\mathbf{y}|\mathbf{X})}{\partial \mathbf{x}_k} = \Phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)\beta_k$$

- The marginal effect is numerically smaller than β_k (in case the relation with $\mathbf{x}_k$ is linear).
  - If $\mathbf{y}^*$ is negative, the marg. effect of $\mathbf{x}_k$ is zero then the average effect of $\mathbf{x}_k$ is less than β_k
- In fact, the marginal effect is $P(\mathbf{y}^* > 0)\beta_k$
- What should we do if $\mathbf{x}_k$ is discrete?

SYDDANSK UNIVERSITET

Marginal effects

For $\mathbf{x}_k$ continuous explanatory variable:

$$\begin{aligned}\frac{\partial E(\mathbf{y}|\mathbf{X},\mathbf{y}>0)}{\partial \mathbf{x}_k} =& \beta_k \left\{1-\lambda\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)\left[\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}+\lambda\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)\right]\right\} \\ =& \beta_k \theta\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)\end{aligned}$$

- $0 < \theta(\mathbf{X}\boldsymbol{\beta}/\sigma) < 1$ is an *adjustment factor*
- Sign of β_k is the sign of the effect
- The marginal effect is again numerically less than β_j
- When we increase $\mathbf{x}_k$, we add (or remove if $\beta_k < 0$) observations with $\mathbf{y}$ close to 0
- $\Rightarrow$ we dampen the effect of $\mathbf{x}_k$ on the average value of $\mathbf{y}$

Marginal effects

Q: If the variable $\mathbf{x}_k$ is discrete, how do you calculate the effect of interest?

If $\mathbf{x}_k$ changes from 3 to 4:

Marginal effects

Q: If the variable $\mathbf{x}_k$ is discrete, how do you calculate the effect of interest?

If $\mathbf{x}_k$ changes from 3 to 4:

$$E(\mathbf{y}|\mathbf{X}, \mathbf{x}_k = 4) - E(\mathbf{y}|\mathbf{X}, \mathbf{x}_k = 3)$$

$$E(\mathbf{y}|\mathbf{X}, \mathbf{x}_k = 4, \mathbf{y} > 0) - E(\mathbf{y}|\mathbf{X}, \mathbf{x}_k = 3, \mathbf{y} > 0)$$

Estimation

- To derive the log-likelihood function, we need the probability (or density) of $y_j = y$ given $\mathbf{x}_j$:
- We need to distinguish between two cases:
  - The probability of $y_j = 0$. This is the discrete part of the distribution of y_j then a strictly positive probability of exactly this outcome
  - The density of y_j for values above 0. The continuous part of the distribution

Estimation

For $y_j = 0$:

$$f(0|\mathbf{x}_j) = P(y_j = 0|\mathbf{X}) = P(y_j^* \leq 0|\mathbf{X}) = 1 - \Phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\sigma}\right)$$

For $y_j > 0$, the cumulative function of y_j

$$F(y_j|\mathbf{x}_j) = P(y_j \leq y_j|\mathbf{x}_j) = P(\mathbf{y}^* \leq y_j|\mathbf{x}_j) = F^*(y_j|\mathbf{x}_j)$$

We know that the density $y_j^*|\mathbf{X} \sim N(\mathbf{x}_j\boldsymbol{\beta}, \sigma^2)$. Then the density function of y_j:

$$f^*(y_j|\mathbf{x}_j) = \frac{1}{\sigma}\phi\left(\frac{y_j - \mathbf{x}_j\boldsymbol{\beta}}{\sigma}\right)$$

Estimation

The log-lik for observation $(\mathbf{x}_j, y_j)$ can then be written as:

$$\begin{aligned}\ell_j(\beta,\sigma^2) =& \mathbf{1}[y_j=0]\log P(y_j=0|\mathbf{x}_j) + \mathbf{1}[y_j>0]\log f^*(y_j|\mathbf{x}_j)\\ =& \mathbf{1}[y_j=0]\log\left[1-\Phi\left(\frac{y_j-\mathbf{x}_j\boldsymbol{\beta}}{\sigma}\right)\right]\\ &+\mathbf{1}[y_j>0]\log\left[\frac{1}{\sigma}\phi\left(\frac{y_j-\mathbf{x}_j\boldsymbol{\beta}}{\sigma}\right)\right]\end{aligned}$$

- The likelihood is maximised wrt $\boldsymbol{\beta}$ and σ.
- No problems of parameter identification, we do not estimate $\boldsymbol{\beta}/\sigma$ like in the probit model.
- This is because we use the probability of $y_j = y_j$, not only $P(y_j > 0|\mathbf{X})$ as in the probit.

Estimation

- Usual properties of the ML estimators: consistency, asymptotic efficiency and asymptotic normality
- The asymptotic variance of the parameter estimates can be estimated using the usual 3 alternatives for ML
- The usual tests (Wald, LR, and LM) can all be used to test linear restrictions
- Same consequences for the model misspecifications

Reporting results

- Parameter estimates
- In a data censoring model, parameter estimates are all we care about.
- In a corner solution model:
  - Partial effects on $E(\mathbf{y}|\mathbf{X})$ and $E(\mathbf{y}|\mathbf{X}, \mathbf{y} > 0)$
  - The former can be compared to OLS estimates
  - As usual, they must be evaluated at a certain value of $\mathbf{X}$? or averaged across observations
  - Must be computed manually in R
  - Remember to distinguish between marginal effects of continuous and discrete explanatory variables

Exercise

- File tobacco.csv.
- Find out how to estimate the tobit model in R (check in the internet)
- Choose variable $share2$ as dependant and estimate the tobit model on age, $nadults$, $nkids$, $nkids2$ and lnx
- Calculate the average marginal effect of age on the estimated value of $\mathbf{y}$
- Calculate the average marginal effect of age on the estimated value of $\mathbf{y}$ for $\mathbf{y} > 0$
- Calculate the average marginal effect of nkids$=0$ to nkids$=2$ on the estimated value of $\mathbf{y}$
- Run the restricted model without $nkids$ and $nkids2$ and compare it with the full model

Specification issues

- Test for heteroskedasticity
- Endogeneity
- Alternative models

Test for heteroskedasticity

As in the probit model:

- Heteroskedasticity and
- Nonnormality
  - The functional forms are completely different
- Endogeneity

All these result in the Tobit $\hat{\beta}$ being an inconsistent estimator β

Test for heteroskedasticity

We assume that $Var(\boldsymbol{\epsilon}|\mathbf{X}) = \sigma^2 \exp(\mathbf{Z}\boldsymbol{\delta})$, where $\mathbf{Z}$ is a subvector of $\mathbf{X}$ with q variables, but does not include the constant.

1. Estimate the restricted tobit model $H_0 : \delta = 0$
2. Find partial derivatives (gradients) of log-lik wrt β, σ^2, and δ (see equations 17.21 and 17.22 in Wooldridge): $\partial\hat{\ell}_j/\partial\boldsymbol{\beta}$, $\partial\hat{\ell}_j/\partial\sigma^2$ and $\partial\hat{\ell}_j/\partial\boldsymbol{\delta} = \hat{\sigma}^2\mathbf{z}_j(\partial\hat{\ell}_j/\partial\sigma^2)$ by evaluating the gradients at $\boldsymbol{\delta} = 0$ (the H_0) and $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ obtained in step 1
3. OLS the unity vector with $\partial\hat{\ell}_j/\partial\boldsymbol{\beta}$, $\partial\hat{\ell}_j/\partial\sigma^2$ and $\partial\hat{\ell}_j/\partial\boldsymbol{\delta}$ (no intercept)
4. The $LM = n * R^2 = n - SSR$ of regression in step 3
5. $LM \sim \chi^2_q$

Test for heteroskedasticity

Intuition in LM test – remember the general idea:

- Estimate model with restrictions imposed (homokedastic model)
- Check whether the likelihood function from the unrestricted model (possible heteroskedastic model) is close to its maximum when evaluated at restricted estimates
- i.e. the score vector (gradients) from the unrestricted model evaluated at the restricted estimates should be close to zero
- i.e. $\partial\hat{\ell}_j/\partial\boldsymbol{\beta}$, $\partial\hat{\ell}_j/\partial\sigma^2$ and $\partial\hat{\ell}_j/\partial\boldsymbol{\delta}$ should not be systematically different from zero
- i.e they cannot explain a unit vector
- If R^2 of the OLS is very large (the score vector is far from zero), then we reject $H_0 : \boldsymbol{\delta} = 0$ and there is heteroskedasticity

Endogeneity

Inconsistency if one of the explanatory variables is correlated with the error term in the latent model.

If y_1 is the variable we model in the Tobit, and:

$$\begin{aligned}\mathbf{y}_1 =& \max(0, \mathbf{z}_1\delta_{11} + \alpha_1\mathbf{y}_2 + \boldsymbol{\epsilon}_1) \\ \mathbf{y}_2 =& \mathbf{z}_1\delta_{21} + \mathbf{z}_2\delta_{22} + \boldsymbol{\nu}_2\end{aligned}$$

Where $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\nu}_2$ are normal but correlated due to e.g. omitted variables:

- $\mathbf{y}_1$ could be labour supply and
- $\mathbf{y}_2$ could be education.
- Both are affected by unobserved ability.

Endogeneity

$$\mathbf{y}_1 = \max(0, \mathbf{z}_1\delta_{11} + \alpha_1\mathbf{y}_2 + \boldsymbol{\epsilon}_1)$$
$$\mathbf{y}_2 = \mathbf{z}_1\delta_{21} + \mathbf{z}_2\delta_{22} + \boldsymbol{\nu}_2$$

Solutions:

- 2-step procedure (Smith–Blundell)
- Or MLE on the full model $f(\mathbf{y}_1, \mathbf{y}_2|\mathbf{z})$
- In both cases, we need instruments, $\mathbf{z}_2$.

Endogeneity

$$\begin{aligned} \mathbf{y}_1 &= \max(0, \mathbf{z}_1\delta_{11} + \alpha_1\mathbf{y}_2 + \boldsymbol{\epsilon}_1) \\ \mathbf{y}_2 &= \mathbf{z}_1\delta_{21} + \mathbf{z}_2\delta_{22} + \boldsymbol{\nu}_2 \end{aligned}$$

2-step procedure:

1. OLS of $\mathbf{y}_2$ on $\mathbf{z}_1$ and $\mathbf{z}_2$ $\Rightarrow$ get residuals $\hat{\boldsymbol{\nu}}_2$,
2. Tobit: $\mathbf{y}_1 = \max(0, \mathbf{z}_1\delta_{11} + \alpha_1\mathbf{y}_2 + \theta_1\hat{\boldsymbol{\nu}}_2 + \boldsymbol{\epsilon}_1)$ $\Rightarrow$ consistent estimates of parameters
3. t-statistic tests $H_0 : \theta_1 = 0$. If p-value is small then we have endogeneity.
4. With endogeneity, standard errors from step 2 need to be corrected (see Wooldridge pp. 660–662 for details)

Alternative models

In the Tobit model, the same function is determining:

- $\mathbf{y} = 0$ vs $\mathbf{y} > 0$, and
- the value of $\mathbf{y}$ given that $\mathbf{y} > 0$

This is a VERY restrictive assumption

If we want to relax this, we can estimate:

$$P(\mathbf{y} = 0|\mathbf{X}) = 1 - \Phi(\mathbf{X}\boldsymbol{\delta})$$
$$\log(\mathbf{y}|\mathbf{X}, \mathbf{y} > 0) \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2)$$

Now, we use two different sets of parameters, $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$

SYDDANSK UNIVERSITET

Alternative models

$$P(\mathbf{y} = 0|\mathbf{X}) = 1 - \Phi(\mathbf{X}\boldsymbol{\delta}) \quad \text{Choose that } \mathbf{y} > 0 \text{ or not}$$
$$\log(\mathbf{y}|\mathbf{X}, \mathbf{y} > 0) \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2) \quad \text{Estimates the quantity for } \mathbf{y} > 0$$

1. Find $\hat{\boldsymbol{\delta}}$ by ML estimation, i.e. Probit on the binary choice: $\mathbf{y} = 0$ vs $\mathbf{y} > 0$
2. OLS of $\log(\mathbf{y})$ on $\mathbf{X}$

Problem: Cannot be tested directly against the Tobit (they are two different models)

Alternative models

But we can find (estimate):

$$\begin{aligned} E(\mathbf{y}|\mathbf{X}, \mathbf{y} > 0) =& \exp(\mathbf{X}\boldsymbol{\beta} + \sigma^2/2) \\ E(\mathbf{y}|\mathbf{X}) =& \Phi(\mathbf{X}\boldsymbol{\delta}) \exp(\mathbf{X}\boldsymbol{\beta} + \sigma^2/2) \end{aligned}$$

And (informally) compare these with those from the Tobit

Summary

- Situation: $\mathbf{y}$ continuous with $P(\mathbf{y} = 0) > 0$
- Two reasons:
  1. Data censoring (coding reason)
  2. Corner solution (behavioural reason)
- In both cases, we assume a latent model: $\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- Only in case 1, are we directly interested in $\boldsymbol{\beta}$; in case 2, we want the marginal effects on the expected means.

Summary

- Estimation:
  - ML
  - Usual properties and tests
- Specification issues:
  - Heterogeneity (test using LM)
  - Endogeneity (two-step procedure)
  - Non-normality
- Alternative model (more general)