#### Econometrics 2 - Lecture 2

## Models with Limited Dependent Variables

#### Contents

- Limited Dependent Variable Cases
- Binary Choice Models
- Binary Choice Models: Estimation
- Binary Choice Models: Goodness of Fit
- Application to Latent Models
- Multi-response Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- The Tobit Model: Estimation
- The Tobit II Model

#### Example: Car Ownership

Explain whether a household owns a car: explanatory power have

- income
- household size
- etc.

Regression for describing car-ownership is not suitable!

- Owning a car has two manifestations: yes/no
- Indicator for owning a car is a binary variable
- Models are needed that allow to describe a binary dependent variable or a, more generally, limited dependent variable

#### Cases of Limited Dependent Variables

- Typical situations: functions of explanatory variables are used to describe or explain
- Dichotomous or binary dependent variable, e.g., ownership of a car (yes/no), employment status (employed/unemployed), etc.
- Ordered response, e.g., qualitative assessment (good/average/bad), working status (full-time/part-time/not working), etc.
- Multinomial response, e.g., trading destinations (Europe/Asia/Africa), transportation means (train/bus/car), etc.
- Count data, e.g., number of orders a company receives in a week, number of patents granted to a company in a year
- Censored data, e.g., expenditures for durable goods, duration of study with drop outs

## Example: Car Ownership and Income

What is the probability that a randomly chosen household owns a car?

- Sample of *N*=32 households, among them 19 households with car
  - Proportion of car owning households:19/32 = 0.59
- Estimated probability for owning a car: 0.59
- But: The probability may differ for rich and poor!
- The sample data contain income information:
  - Yearly income: average EUR 20.524, minimum EUR 12.000, maximum EUR 32.517
  - Proportion of car owning households among the 16 households with less than EUR 20.000 income: 9/16 = 0.56
  - Proportion of car owning households among the 16 households with more than EUR 20.000 income: 10/16 = 0.63

#### Car Ownership and Income, cont'd

How can a model for the probability – or prediction – of car ownership take the income of a household into account?

Notation: N households

- dummy  $y_i$  for car ownership;  $y_i = 1$ : household *i* has car
- income of *i*-th household:  $x_{i2}$
- For predicting  $y_i$  or estimating the probability  $P\{y_i = 1\}$  , a model is needed that takes the income into account

## Modelling Car Ownership

How is car ownership related to the income of a household?

- 1. Linear regression  $y_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i = x_i'\beta + \varepsilon_i$
- With  $E{\epsilon_i | x_i} = 0$ , the model  $y_i = x_i'\beta + \epsilon_i$  gives

 $\mathsf{P}\{y_i = 1 | x_i\} = x_i'\beta$ 

due to  $E\{y_i|x_i\} = x_i'\beta = 1*P\{y_i = 1|x_i\} + 0*P\{y_i = 0|x_i\} = P\{y_i = 1|x_i\}$ 

- The systematic part of  $y_i = x_i'\beta + \varepsilon_i$ ,  $x_i'\beta$ , is  $P\{y_i = 1 | x_i\}!$
- Model for y is specifying the probability for y = 1 as a function of x

#### Problems:

- $x_i'\beta$  not necessarily in [0,1]
- Error terms: for a given  $x_i$ 
  - $\epsilon_i$  can take on only two values, viz. 1-  $x_i'\beta$  and  $x_i'\beta$
  - $V{\epsilon_i | x_i} = x_i'\beta(1 x_i'\beta)$ , heteroskedastic, dependent upon  $\beta$

#### Modelling Car Ownership, cont'd

2. Use of a function  $G(x_i,\beta)$  with values in the interval [0,1]  $P\{y_i = 1 | x_i\} = E\{y_i | x_i\} = G(x_i, \beta)$ Standard logistic distribution function  $L(z) = \frac{e^{z}}{1 + e^{z}} = \frac{1}{1 + e^{-z}}$ L(z) fulfils  $\lim_{z\to -\infty} L(z) = 0$ ,  $\lim_{z\to \infty} L(z) = 1$ Binary choice model:  $P\{y_i = 1 | x_i\} = p_i = L(x_i'\beta) = [1 + \exp\{-x_i'\beta\}]^{-1}$ Can be written using the odds ratio  $p_i/(1 - p_i)$  for the event  $\{y_i = 1 | x_i\}$  $\log \frac{p_i}{1 - p_i} = x_i' \beta$ Interpretation of coefficients  $\beta$ : An increase of  $x_{i2}$  by 1 results in a relative change of the odds ratio  $p_i/(1-p_i)$  by  $\beta_2$  or by 100 $\beta_2$ %; cf. the

#### Car Ownership and Income, cont'd

E	E.g., $P{y_i = 1 x_i} = 1/(1+exp(-z_i))$ with $z = -0.5 + 1.1*x$ , the income x in EUR 1000 per month								
•	Increasing income is associated with an increasing probability of owning a car: z goes up by 1.1 for every additional EUR 1000								
•	For a person with an income of EUR 1000, z = 0.6 and the probability of owning a car is 1/(1+exp(-0.6)) = 0.646								
S	Standard	logi	stic d	istribution fu	unction <i>L</i> ( <i>z</i> ), with <i>z</i> on the horizontal				
	and <i>L</i> ( <i>z</i> ) on the vertical axis								
	Income x z $P{y=1 x}$								
	1000	1	0.6	0.646	0.5				
	2000	2	1.7	0.846					
	3000	3	2.8	0.943					

## Odds, Odds Ratio

The odds or the odds ratio (in favour) of event A is the ratio of the probability that A will happen to the probability that A will not happen

- If the probability of success is 0.8 (that of failure is 0.2), the odds of success are 0.8/0.2 = 4; we say, "the odds of success are 4 to 1"
- If the probability of event A is p, that of "not A" therefore being 1-p, the odds or the odds ratio of event A is the ratio p/(1-p)
- We say the odds (ratio) of A is "p/(1-p) to 1" or "1 to (1-p)/p"

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
p/(1-p)	0.11	0.25	0.43	0.67	1	1.5	2.33	4	9
odds	1:9	1:4	1:2.3	1:1.5	1:1	1:0.67	1:0.43	1:0.25	1:0.11

• The logarithm of the odds,  $\log p/(1-p)$ , is called the logit of p

#### Betting Odds

- The probability of success is 0.8
- The odds of success are 4 to 1
- Betting odds for success are 1:4
  - The bookmaker is prepared to pay out a prize of one fourth of the stake and return the stake as well, to anyone who places a bet on success

#### Contents

- Limited Dependent Variable Cases
- Binary Choice Models
- Binary Choice Models: Estimation
- Binary Choice Models: Goodness of Fit
- Application to Latent Models
- Multi-response Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- The Tobit Model: Estimation
- The Tobit II Model

#### **Binary Choice Models**

Model for probability P{ $y_i = 1 | x_i$ }, function of *K* (numerical or categorical) explanatory variables  $x_i$  and unknown parameters  $\beta$ , such as E{ $y_i | x_i$ } = P{ $y_i = 1 | x_i$ } = G( $x_i, \beta$ )

Typical functions  $G(x_i,\beta)$ : distribution functions (cdf's)  $F(x_i'\beta) = F(z)$ 

Probit model: standard normal distribution function; V{z} = 1

$$F(z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^{2}) dt$$

• Logit model: standard logistic distribution function;  $V{z}=\pi^2/3=1.81^2$ 

$$F(z) = L(z) = \frac{e^{z}}{1 + e^{z}}$$

Linear probability model (LPM)

$$f(z) = 0, \, z < 0$$

F

$$= z, 0 \le z \le 1$$

## Linear Probability Model (LPM)

Assumes that

 $\mathsf{P}\{y_i = 1 | x_i\} = x_i'\beta \text{ for } 0 \le x_i'\beta \le 1$ 

but sets restrictions

 $P\{y_i = 1 | x_i\} = 0 \text{ for } x_i'\beta < 0$  $P\{y_i = 1 | x_i\} = 1 \text{ for } x_i'\beta > 1$ 

- Typically, the model is estimated by OLS, ignoring the probability restrictions
- Standard errors should be adjusted using heteroskedasticityconsistent (White) standard errors

#### Probit Model: Standardization

 $E{y_i|x_i} = P{y_i = 1|x_i} = F(x_i'\beta)$ : assume *F*(.) to be the distribution function of the standard normal distribution N(0, σ<sup>2</sup>)

$$P\left\{y_i = 1 \mid x_i\right\} = \Phi\left(\frac{x_i \cdot \beta}{\sigma}\right)$$

- Given  $x_i$ , the ratio  $\beta/\sigma^2$  determines  $P\{y_i = 1 | x_i\}$
- Standardization restriction  $\sigma^2 = 1$ : allows unique estimates for  $\beta$

#### Probit vs Logit Model

- Differences between the probit and the logit model:
  - Shapes of distribution are slightly different, particularly in the tails
  - Scaling of the distributions is different: The implicit variance for  $\varepsilon_i$  in the logit model is  $\pi^2/3 = (1.81)^2$ , while 1 for the probit model
  - Probit model is relatively easy to extend to multivariate cases using the multivariate normal or conditional normal distribution
- In practice, the probit and logit model produce quite similar results
  - The scaling difference makes the values of  $\beta$  not directly comparable across the two models, while the signs are typically the same
  - □ The estimates of β in the logit model are roughly a factor  $π/\sqrt{3} \approx 1.81$  larger than those in the probit model

#### Marginal Effects of Binary Choice Models

Linear regression model  $E\{y_i|x_i\} = x_i'\beta$ : the marginal effect  $\partial E\{y_i|x_i\}/\partial x_{ik}$ of a change in  $x_k$  is  $\beta_k$ 

For  $E\{y_i|x_i\} = F(x_i'\beta)$ 

$$\frac{\partial E\{y_i \mid x_i\}}{\partial x_k} = \frac{\partial F(x_i \mid \beta)}{\partial x_k} \beta_k$$

• The marginal effect of changing  $x_k$ 

Probit model:  $\phi(x_i'\beta) \beta_k$ , with standard normal density function  $\phi$ 

□ Logit model:  $\exp\{x_i'\beta\}/[1 + \exp\{x_i'\beta\}]^2\beta_k$ 

Linear probability model:  $\beta_k$  if  $x_i'\beta$  is in [0,1]

In general, the marginal effect of changing the regressor  $x_k$  depends upon  $x_i$ ' $\beta$ , the shape of F, and  $\beta_k$ ; the sign is that of  $\beta_k$ 

#### Interpretation of Binary Choice Models

The effect of a change in  $x_k$  can be characterized by the

 "Slope", i.e., the "average" marginal effect or the gradient of E{y<sub>i</sub>|x<sub>i</sub>} for the sample means of the regressors

$$slope_k(\overline{x}) = \frac{\partial F(x_i' \beta)}{\partial x_k} \bigg|_{\overline{x}}$$

- For a dummy variable *D*: marginal effect is calculated as the difference of probabilities  $P\{y_i = 1 | x_{(d)}, D = 1\} P\{y_i = 1 | x_{(d)}, D = 0\}; x_{(d)}$  stands for the sample means of all regressors except *D*
- For the logit model:

$$\log \frac{p_i}{1 - p_i} = x_i \, '\beta$$

The coefficient  $\beta_k$  is the relative change of the odds ratio when increasing  $x_k$  by 1 unit

#### Contents

- Limited Dependent Variable Cases
- Binary Choice Models
- Binary Choice Models: Estimation
- Binary Choice Models: Goodness of Fit
- Application to Latent Models
- Multi-response Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- The Tobit Model: Estimation
- The Tobit II Model

#### Binary Choice Models: Estimation

Typically, binary choice models are estimated by maximum likelihood Likelihood function, given *N* observations ( $y_i$ ,  $x_i$ )

$$L(\beta) = \prod_{i=1}^{N} P\{y_i = 1 | x_i; \beta\}^{y_i} P\{y_i = 0 | x_i; \beta\}^{1-y_i}$$

= 
$$\prod_{i} F(x_{i}'\beta)^{y_{i}} (1 - F(x_{i}'\beta))^{1-y_{i}}$$

Maximization of the log-likelihood function

 $\ell(\beta) = \log L(\beta) = \sum_{i} y_{i} \log F(x_{i}'\beta) + \sum_{i} (1-y_{i}) \log (1-F(x_{i}'\beta))$ 

First-order conditions of the maximization problem

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i} \left[ \frac{y_i - F(x_i \, \beta)}{F(x_i \, \beta)(1 - F(x_i \, \beta))} f(x_i \, \beta) \right] x_i = \sum_{i} e_i x_i = 0 \quad (\mathsf{A})$$

e<sub>i</sub>: generalized residuals

#### Generalized Residuals

The first-order conditions  $\Sigma_i e_i x_i = 0$  define the generalized residuals  $e_i = \frac{y_i - F(x_i \cdot \beta)}{F(x_i \cdot \beta)(1 - F(x_i \cdot \beta))} f(x_i \cdot \beta)$ 

• The generalized residuals  $e_i$  can assume two values, depending on the value of  $y_i$ :

$$\Box \quad e_i = f(x_i'b)/F(x_i'b) \text{ if } y_i = 1$$

$$= e_i = -f(x_i'b)/(1-F(x_i'b)) \text{ if } y_i = 0$$

*b* are the estimates of  $\beta$ 

 Generalized residuals are orthogonal to each regressor; cf. the first-order conditions of OLS estimation

#### Estimation of Logit Model

First-order condition of the maximization problem

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i} \left[ y_{i} - \frac{\exp(x_{i} \mid \beta)}{1 + \exp(x_{i} \mid \beta)} \right] x_{i} = 0$$
  
gives [due to P{ $y_{i} = 1 \mid x_{i}$ } =  $p_{i} = L(x_{i}, \beta)$ ]  
 $\hat{p}_{i} = \frac{\exp(x_{i} \mid b)}{1 + \exp(x_{i} \mid b)}$ 

- From  $\Sigma_i \hat{p}_i x_i = \Sigma_i y_i x_i$  follows given that the model contains an intercept –:
  - $\square$  The sum of estimated probabilities  $\Sigma_{i} \hat{p}_{i}~$  equals the observed frequency  $\Sigma_{i} y_{i}$
- Similar results for the probit model, due to similarity of logit and probit functions

## Binary Choice Models in GRETL

Model > Limited dependent variable > Logit > Binary

 Estimates the specified model using error terms with standard logistic distribution

Model > Limited dependent variable > Probit > Binary

 Estimates the specified model using error terms with standard normal distribution

## Example: Effect of Teaching Method

- Study by Spector & Mazzeo (1980); Greene's dataset "greene19\_1" (Data on educational program effectiveness)
- Personalized System of Instruction: has the new teaching method in economics an effect on student performance in later courses?

#### Data:

- GRADE (0/1): indicator whether grade was higher than in principal course
- PSI (0/1): participation in program with new teaching method
- GPA: grade point average
- TUCE: score on a pre-test, entering knowledge
- 32 observations

	mean	min	max
GPA	3.12	2.06	4.00
TUCE	21.9	12	29

#### Effect of Teaching Method, cont'd

#### Logit model for GRADE, GRETL output

Model 1: Logit, using observations 1-32
Dependent variable: GRADE

const	Coefficient	Std. Error	<i>z-stat</i> -2.6405	Slope*	
GPA	2 82611	1 26294	2 2377	0 533859	
TUCE	0.0951577	0.141554	0.6722	0.0179755	
PSI	2.37869	1.06456	2.2344	0.456498	
Mean deper	ndent var	0.343750		S.D. dependent var	0.188902
McFadden F	R-squared	0.374038		Adjusted R-squared	0.179786
Log-likelihoo	bd	-12.88963		Akaike criterion	33.77927
Schwarz crit	terion	39.64221		Hannan-Quinn	35.72267
*Number of o f(beta'x) at n Likelihood ra	cases 'correctly pr nean of independe atio test: Chi-squa	edicted' = 26 (8 ent vars = 0.189 re(3) = 15.4042	1.3%) [0.0015]		
	Predicted				
Actual	0 18 3				
Actual	1 3 8				

### Effect of Teaching Method, cont'd

Estimated logit model for the indicator GRADE  $P{GRADE = 1} = p = L(z) = \exp{z}/(1+\exp{z})$ with

*z* = -13.02 + 2.826\*GPA + 0.095\*TUCE + 2.38\*PSI

$$= \log \{p/(1-p)\} = \log \{p\}$$

- Regressors
  - GPA: grade point average
  - TUCE: score on a pre-test, entering knowledge
  - PSI (0/1): participation in program with new teaching method
- Slopes
  - **GPA: 0.53**
  - **TUCE: 0.02**
  - Difference P{GRADE =1 $|x_{(d)}$ , PSI=1} P{GRADE =1 $|x_{(d)}$ , PSI=0}: 0.49; cf. Slope 0.46



#### **Properties of ML Estimators**

#### Consistent

- Asymptotically efficient
- Asymptotically normally distributed

These properties require that the assumed distribution is correct

- Correct shape
- No autocorrelation and/or heteroskedasticity
- No dependence correlations between errors and regressors
- No omitted regressors

#### Contents

- Limited Dependent Variable Cases
- Binary Choice Models
- Binary Choice Models: Estimation
- Binary Choice Models: Goodness of Fit
- Application to Latent Models
- Multi-response Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- The Tobit Model: Estimation
- The Tobit II Model

#### Goodness-of-fit Measures

#### Concepts

- Comparison of the maximum likelihood of the model with that of the naïve model, i.e., a model with only an intercept, no regressors
  - □ *pseudo*-R<sup>2</sup>
  - McFadden R<sup>2</sup>
- Goodness-of-fit measure based on proportion of correctly predicted observations or hit rates

$$\square R_p^2$$

#### *McFadden* R<sup>2</sup>

Based on log-likelihood function

- $l(b) = l_1$ : maximum log-likelihood of the model to be assessed
- $\ell_0$ : maximum log-likelihood of the naïve model, i.e., a model with only an intercept;  $\ell_0 \le \ell_1$  and  $\ell_0$ ,  $\ell_1 < 0$ 
  - The larger  $l_1 l_0$ , the more contribute the regressors
  - $l_1 = l_0, if all slope coefficients are zero$
  - $larrow l_1 = 0$ , if  $y_i$  is exactly predicted for all *i*
- pseudo-R<sup>2</sup>: a number in [0,1), defined by

$$pseudo - R^{2} = 1 - \frac{1}{1 + 2(\ell_{1} - \ell_{0}) / N}$$

- *McFadden* R<sup>2</sup>: a number in [0,1], defined by  $McFadden R^2 = 1 \ell_1 / \ell_0$
- Both are 0 if  $l_1 = l_0$ , i.e., all slope coefficients are zero
- *McFadden*  $\mathbb{R}^2$  attains the upper limit 1 if  $\ell_1 = 0$

## Naïve Model: Calculation of $\ell_0$

Maximum log-likelihood function of the naïve model, i.e., a model with only an intercept:  $l_0$ 

- Not necessary to estimate a logit or probit model
- $P{y_i = 1} = p$  for all *i* (cf. urn experiment)
- Log-likelihood function

 $\log L(p) = N_1 \log(p) + (N - N_1) \log (1-p)$ 

with  $N_1 = \Sigma_i y_i$ , i.e., the observed frequency

- Maximum likelihood estimator for p is  $N_1/N$
- Maximum log-likelihood of the naïve model

 $\ell_0 = N_1 \log(N_1/N) + (N - N_1) \log (1 - N_1/N)$ 

## Goodness-of-fit Measure R<sub>p</sub><sup>2</sup>

Comparison of correct and incorrect predictions

Predicted outcome

$$\hat{y}_i = 1$$
 if  $F(x_i'b) > 0.5$ , i.e., if  $x_i'b > 0$ 

= 0 if 
$$F(x_i'b) \le 0.5$$
, i.e., if  $x_i'b \le 0$ 

- Cross-tabulation of actual and predicted outcome
- Proportion of incorrect predictions  $wr_1 = (n_{01} + n_{10})/N$
- Hit rate:  $1 wr_1$

proportion of correct predictions

- Comparison with naive model:
  - Predicted outcome of naïve model

 $\hat{y}_i = 1$  for all *i* (!), if  $\hat{p} = N_1/N > 0.5$ ;  $\hat{y}_i = 0$  for all *i* if  $\hat{p} \le 0.5$ 

• 
$$wr_0 = 1 - \hat{p}$$
 if  $\hat{p} > 0.5$ ,  $wr_0 = \hat{p}$  if  $\hat{p} \le 0.5$ 

• Goodness-of-fit measure:  $R_p^2 = 1 - wr_1/wr_0$ ; may be negative!

	ŷ = 0	ŷ = 1	Σ
<i>y</i> = 0	<i>n</i> <sub>00</sub>	n <sub>01</sub>	$N_0$
<i>y</i> = 1	n <sub>10</sub>	n <sub>11</sub>	<i>N</i> <sub>1</sub>
Σ	$n_0$	n <sub>1</sub>	Ν

#### Effect of Teaching Method, cont'd

Logit model for	GRADE,	GRETL	output
-----------------	--------	-------	--------

const	Coefficient -13.0213	<i>Std. Error</i> 4.93132	<i>z-stat</i> -2.6405	Slope*	
GPA TUCE	2.82611 0.0951577	1.26294 0 141554	2.2377	0.533859 0.0179755	
PSI	2.37869	1.06456	2.2344	0.456498	
Mean depe	endent var	0.343750		S.D. dependent var	0.18890
McFadden	R-squared	0.374038		Adjusted R-squared	0.179786
Log-likeliho	bod	-12.88963		Akaike criterion	33.7792
Schwarz cr	iterion	39.64221		Hannan-Quinn	35.722
Number of	f cases 'correctly pr	edicted' = 26 (8	1.3%)		
Likelihood	ratio test: Chi-squa	re(3) = 15.4042	[0.0015]		
	Predicted				



# Effect of Teaching Method, cont'd

Comparison of the LPM, logit, and probit model for GRADE
Estimated models: coefficients and their standard errors

	LPM		Logit		Probit	
	coeff	slope	coeff	slope	coeff	slope
const	-1.498		-13.02		-7.452	
GPA	0.464	0.464	2.826	0.534	1.626	0.533
TUCE	0.010	0.010	0.095	0.018	0.052	0.017
PSI	0.379	0.379	2.379	0.456	1.426	0.464

- Coefficients of logit model: due to larger variance, larger by factor  $\sqrt{(\pi^2/3)}=1.81$  than that of the probit model
- Very similar slopes
# Effect of Teaching Method, cont'd

Goodness-of-fit measures for the logit model

• With  $N_1 = 11$  and N = 32

 $\ell_0 = 11 \log(11/32) + 21 \log(21/32) = -20.59$ 

• As  $\hat{p} = N_1/N = 0.34 < 0.5$ : the proportion  $wr_0$  of incorrect predictions with the naïve model is

$$wr_0 = \hat{p} = 11/32 = 0.34$$

From the GRETL output:  $l_1 = -12.89$ ,  $wr_1 = 6/32$ 

Goodness-of-fit measures

- McFadden R<sup>2</sup> = 1 (-12.89)/(-20.59) = 0.374
- $pseudo-R^2 = 1 1/(1 + 2(-12.89 + 20.59)/32) = 0.325$

$$R_{0}^{2} = 1 - wr_{1}/wr_{0} = 1 - 6/11 = 0.45$$

#### Contents

- Limited Dependent Variable Cases
- Binary Choice Models
- Binary Choice Models: Estimation
- Binary Choice Models: Goodness of Fit
- Application to Latent Models
- Multi-response Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- The Tobit Model: Estimation
- The Tobit II Model

# Modelling Utility

Latent variable y<sub>i</sub>\*: utility difference between owning and not owning a car; unobservable (latent)

- Decision on owning a car
  - $y_i^* > 0$ : in favour of car owning
  - □  $y_i^* \le 0$ : against car owning
- y<sub>i</sub>\* depends upon observed characteristics (e.g., income) and unobserved characteristics ε<sub>i</sub>

 $y_i^* = x_i'\beta + \varepsilon_i$ 

• Observation  $y_i = 1$  (i.e., owning car) if  $y_i^* > 0$ 

 $\mathsf{P}\{y_i = 1\} = \mathsf{P}\{y_i^* > 0\} = \mathsf{P}\{x_i'\beta + \varepsilon_i > 0\} = 1 - F(-x_i'\beta) = F(x_i'\beta)$ 

last step requires a distribution function F(.) with symmetric density Latent variable model: based on a latent variable that represents the underlying behaviour

#### Latent Variable Model

Model for the latent variable  $y_i^*$ 

 $y_i^* = x_i'\beta + \varepsilon_i$ 

 $y_i^*$ : not necessarily a utility difference

- $\epsilon_i$ 's are independent of  $x_i$ 's
- ε<sub>i</sub> has a standardized distribution
  - Probit model if  $\varepsilon_i$  has standard normal distribution
  - Logit model if  $\varepsilon_i$  has standard logistic distribution
- Observations
  - $y_i = 1$  if  $y_i^* > 0$
  - $\Box \quad y_i = 0 \text{ if } y_i^* \le 0$
- ML estimation

#### Contents

- Limited Dependent Variable Cases
- Binary Choice Models
- Binary Choice Models: Estimation
- Binary Choice Models: Goodness of Fit
- Application to Latent Models
- Multi-response Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- The Tobit Model: Estimation
- The Tobit II Model

#### Multi-response Models

Models for explaining the choice between discrete outcomes

- Examples:
  - a. Working status (full-time/part-time/not working), qualitative assessment (good/average/bad), etc.
  - b. Trading destinations (Europe/Asia/Africa), transportation means (train/bus/car), etc.
- Multi-response models describe the probability of each of these outcomes, as a function of variables like
  - person-specific characteristics
  - alternative-specific characteristics
- Types of multi-response models (cf. above examples)
  - Ordered response models: outcomes have a natural ordering
  - Multinomial (unordered) models: ordering of outcomes is arbitrary

### Example: Credit Rating

Credit rating: numbers, indicating experts' opinion about (a firm's) capacity to satisfy financial obligations, e.g., credit-worthiness

- Standard & Poor's rating scale: AAA, AA+, AA, AA-, A+, A, A-, BBB+, BBB, BBB-, BB+, BB, BB-, B+, B, B-, CCC+, CCC, CCC-, CC, C, D
- Verbeek's data set CREDIT
  - □ Categories "1", ..., "7" (highest)
  - Investment grade with alternatives "1" (better than category 3) and "0" (category 3 or less, also called "speculative grade")
- Explanatory variables, e.g.,
  - Firm sales
  - Ebit, i.e., earnings before interest and taxes
  - Ratio of working capital to total assets

#### Ordered Response Model

Choice between *M* alternatives

Observed alternative for sample unit *i*:  $y_i$ 

Latent variable model

 $y_i^* = x_i'\beta + \varepsilon_i$ 

with K-vector of explanatory variables  $x_i$ 

 $y_i = j$  if  $\gamma_{i-1} < y_i^* \le \gamma_i$  for  $j = 0, \dots, M$ 

- *M*+1 boundaries  $\gamma_i$ , j = 0, ..., M, with  $\gamma_0 = -\infty, ..., \gamma_M = \infty$
- $ε_i$ 's are independent of  $x_i$ 's
- ε<sub>i</sub> typically follows the
  - standard normal distribution: ordered probit model
  - standard logistic distribution: ordered logit model

# Example: Willingness to Work

Married females are asked: "How much would you like to work?"

- Potential answers of individual *i*:  $y_i = 1$  (not working),  $y_i = 2$  (part time),  $y_i = 3$  (full time)
- Measure of the desired labour supply
- Dependent upon factors like age, education level, husband's income Ordered response model with M = 3

$$y_i^* = x_i'\beta + \varepsilon_i$$

with

$$y_i = 1 \text{ if } y_i^* \le 0$$
  

$$y_i = 2 \text{ if } 0 < y_i^* \le \gamma$$
  

$$v_i = 3 \text{ if } v_i^* > \gamma$$

- $\epsilon_{i}$ 's with distribution function F(.)
  - y<sub>i</sub>\* stands for "willingness to work" or "desired hours of work"

#### Willingness to Work, cont'd

In terms of observed quantities:

$$P\{y_{i} = 1 | x_{i}\} = P\{y_{i}^{*} \le 0 | x_{i}\} = F(-x_{i}'\beta)$$

$$P\{y_{i} = 3 | x_{i}\} = P\{y_{i}^{*} > \gamma | x_{i}\} = 1 - F(\gamma - x_{i}'\beta)$$

$$P\{y_{i} = 2 | x_{i}\} = F(\gamma - x_{i}'\beta) - F(-x_{i}'\beta)$$

- Unknown parameters: γ and β
- Standardization: wrt location (boundary 0) and scale (V{ $\epsilon_i$ } = 1)
- ML estimation

#### Interpretation of parameters β

- Wrt y<sub>i</sub>\*(= x<sub>i</sub>'β + ε<sub>i</sub>): willingness to work increases with larger x<sub>k</sub> for positive β<sub>k</sub>
- Wrt probabilities  $P\{y_i = j | x_i\}$ , e.g., for positive  $\beta_k$ 
  - $\square P\{y_i = 3 | x_i\} = P\{y_i^* > \gamma | x_i\} \text{ increases and}$
  - □  $P{y_i = 1 | x_i} P{y_i^* \le 0 | x_i}$  decreases with larger  $x_k$

### Example: Credit Rating

Verbeek's data set CREDIT: 921 observations for US firms' credit ratings in 2005, including firm characteristics Rating models:

- Ordered logit model for assignment of categories "1", ..., "7" (highest)
- Binary logit model for assignment of "investment grade" with alternatives "1" (better than category 3) and "0" (category 3 or less, also called "speculative grade")

#### Credit Rating, cont'd

#### Verbeek's data set CREDIT

Ratings and characteristics for 921 firms: summary statistics

Table 7.4	Summary statistics				
	average	median	minimum	maximum	
credit rating	3.499	3	$\begin{array}{c} 1 \\ 0 \end{array}$	7	
investment grade	0.472	0		1	
book leverage	0.293	0.264	$0.000 \\ -0.412$	0.999	
working capital/total assets	0.140	0.123		0.748	
retained earnings/total assets	0.157	0.180	$-0.996 \\ -0.384 \\ 1.100$	0.980	
earnings before interest and taxes/t.a	. 0.094	0.090		0.652	
log sales	7.996	7.884		12.701	

#### Book leverage: ratio of debts to total assets

#### Credit Rating, cont'd

#### Verbeek, Table 7.5.

	Binary logit			Ordered logit		
	Estimate	Standard error		Estimate	Standard erro	
constant	-8.214	0.867		_	_	
book leverage	-4.427	0.771		-2.752	0.477	
ebit/ta	4.355	1.440		4.731	0.945	
log <i>sales</i>	1.082	0.096		0.941	0.059	
re/ta	4.116	0.489		3.560	0.302	
wk/ta	-4.012	0.748		-2.580	0.483	
			$\gamma_1$	-0.369	0.633	
			$\gamma_2$	4.881	0.521	
			$\gamma_3$	7.626	0.551	
			$\gamma_4$	9.885	0.592	
			$\gamma_5$	12.883	0.673	
			$\gamma_6$	14.783	0.784	
loglikelihood	-341.08			-965.31		
McFadden $R^2$	0.465			0.309		
LR test $(\chi_{\epsilon}^2)$	$591.8 \ (p = 0.000)$			$862.9 \ (p = 0.000)$		

#### Ordered Response Model: Estimation

Latent variable model

 $y_i^* = x_i'\beta + \varepsilon_i$ 

with explanatory variables  $x_i$ 

 $y_i = j$  if  $\gamma_{j-1} < y_i^* \le \gamma_j$  for  $j = 0, \dots, M$ 

ML estimation of  $\beta_1,\,...,\,\beta_K$  and  $\gamma_1,\,...,\,\gamma_{M\text{-}1}$ 

- Log-likelihood function in terms of probabilities
- Numerical optimization
- ML estimators are
  - Consistent
  - Asymptotically efficient
  - Asymptotically normally distributed

#### Contents

- Limited Dependent Variable Cases
- Binary Choice Models
- Binary Choice Models: Estimation
- Binary Choice Models: Goodness of Fit
- Application to Latent Models
- Multi-response Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- The Tobit Model: Estimation
- The Tobit II Model

#### Multinomial Models

Choice between *M* alternatives without natural order Observed alternative for sample unit *i*:  $y_i$ 

"Random utility" framework: Individual i

- attaches utility levels  $U_{ij}$  to each of the alternatives, j = 1, ..., M,
- chooses the alternative with the highest utility level max{ $U_{i1}, ..., U_{iM}$ } Utility levels  $U_{ij}, j = 1, ..., M$ , as a function of characteristics  $x_{ij}$  $U_{ij} = x_{ij}'\beta + \varepsilon_{ij} = \mu_{ij} + \varepsilon_{ij}$
- error terms  $\varepsilon_{ij}$  follow the Type I extreme value distribution: leads to  $P\{y_i = j\} = \frac{\exp\{\mu_{ij}\}}{\exp\{\mu_{i1}\} + \dots + \exp\{\mu_{iM}\}} = \frac{\exp\{x_{ij} \mid \beta\}}{\exp\{x_{i1} \mid \beta\} + \dots + \exp\{x_{iM} \mid \beta\}}$ for  $j = 1, \dots, M$
- and  $\Sigma_j P\{y_i = j\} = 1$

For setting the location: constraint  $x_{i1}$ ' $\beta = \mu_{i1} = 0$  or exp{ $\mu_{i1}$ } = 1

#### Variants of the Logit Model

Conditional logit model: for j = 1, ..., M

$$P\{y_i = j\} = \frac{\exp\{x_{ij} \,' \beta\}}{1 + \exp\{x_{i2} \,' \beta\} + \dots + \exp\{x_{iM} \,' \beta\}}$$

- Alternative-specific characteristics x<sub>ii</sub>
- E.g., mode of transportation (by car, train, bus) is affected by the travel costs, travel time, etc. of the individual i

Multinomial logit model: for j = 1, ..., M

$$P\{y_i = j\} = \frac{\exp\{x_i' \beta_j\}}{1 + \exp\{x_i' \beta_2\} + \dots + \exp\{x_i' \beta_M\}}$$

- Person-specific characteristics x<sub>i</sub>
- E.g., mode of transportation is affected by income, gender, etc.

# Multinomial Logit Model

The term "multinomial logit model" is also used for both the

- the conditional logit model
- the multinomial logit model (see above)
- and also for the mixed logit model: it combines
  - alternative-specific characteristics and
  - person-specific characteristics

Number of parameters

- conditional logit model: vector  $\beta$  with K components
- multinomial logit model: vectors  $\beta_2, ..., \beta_M$ , each with K components

#### Independence of Errors

Independence of the error terms  $\boldsymbol{\epsilon}_{ij}$  implies independent utility levels of alternatives

- Independence assumption may be restrictive
- Example: High utility of alternative "travel with red bus" implies high utility of "travel with blue bus"
- Implies that the odds ratio of two alternatives does not depend upon other alternatives: "independence of irrelevant alternatives" (IIA)

### Multi-response Models in GRETL

Model > Limited dependent variable > Logit > Ordered

- Estimates the specified model using error terms with standard logistic distribution, assuming ordered alternatives for responses
   Model > Limited dependent variable > Logit > Multinomial
- Estimates the specified model using error terms with standard logistic distribution, assuming alternatives without order

```
Model > Limited dependent variable > Probit > Ordered
```

 Estimates the specified model using error terms with standard normal distribution, assuming ordered alternatives

#### Contents

- Limited Dependent Variable Cases
- Binary Choice Models
- Binary Choice Models: Estimation
- Binary Choice Models: Goodness of Fit
- Application to Latent Models
- Multi-response Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- The Tobit Model: Estimation
- The Tobit II Model

#### Models for Count Data

Describe the number of times an event occurs, depending upon certain characteristics

Examples:

- Number of visits in the library per week
- Number of visits of a customer in the supermarket
- Number of misspellings in an email
- Number of applications of a firm for patents, as a function of
  - firm size
  - R&D expenditures
  - industrial sector
  - country, etc.

See Verbeek's data set PATENTS

#### Example: Patents and R&D Expenditures

Verbeek's data set PATENTS: number of patents (p91), expenditures for R&D (log\_rd91), sector of industry, and region; N = 181 Question: Is the number of patents depending of R&D expenditures, sector, region?

#### Poisson Regression Model

Observed variable for sample unit *i*:

*y*<sub>i</sub>: number of possible outcomes 0, 1, ..., *y*, ... Aim: to explain E{*y*<sub>i</sub> | *x*<sub>i</sub>}, based on characteristics *x*<sub>i</sub> E{*y*<sub>i</sub> | *x*<sub>i</sub>} = exp{*x*<sub>i</sub>'β} Poisson regression model  $P\{y_i = y | x_i\} = \frac{\lambda_i^y}{y!} \exp{\{\lambda_i\}}, y = 0, 1, ...$ 

with  $\lambda_i = E\{y_i \mid x_i\} = \exp\{x_i'\beta\}$  $y! = 1 \times 2 \times \dots \times y, 0! = 1$ 

#### **Poisson Distribution**

$$P\{X=k\} = \frac{\lambda^k}{k!} \exp\{\lambda\}, k = 0, 1, ...$$



#### Poisson Regression Model: Estimation

Unknown parameters: coefficients β

Estimates of  $\beta$  allow assessing how exp{ $x_i'\beta$ } = E{ $y_i | x_i$ } is affected by  $x_i$ Fitting the model to data: ML estimators for  $\beta$  are

- Consistent
- Asymptotically efficient
- Asymptotically normally distributed

#### Patents and R&D Expenditures

Verbeek's data set PATENTS: number of patents (p91), log R&D expenditures (Ir91), sector of industry, and region; N = 181Question: Is the number of patents depending of R&D expenditures, sector, region?

Model:

 $\mathsf{E}\{y_i \mid x_i\} = \exp\{x_i'\beta\}$ 

- y<sub>i</sub>: number of patents in company *i* in year 1991
- x<sub>i</sub>: characteristics of company *i*: intercept, R&D expenditures in1991, dummy for sector (aerosp, chemist, computer, machines, vehicles), region (US, Europe, Japan)

```
Variable p91: mean: 73.6, std.dev.: 150.9
Overdispersion ?
```

#### Patents and R&D Expenditures

#### Poisson regression model for p91, GRETL output

Convergence achieved after 8 iterations

Model 1: Poisson, using observations 1-181 Dependent variable: p91

	coefficient	std. error	Z	p-value	
const	-0.873731	0.0658703	-13.26	3.72e-040 ***	
log_rd91	0.854525	0.00838674	101.9	0.0000 ***	
aerosp	-1.42185	0.0956448	-14.87	5.48e-050 ***	
chemist	0.636267	0.0255274	24.92	4.00e-137 ***	
computer	0.595343	0.0233387	25.51	1.57e-143 ***	
machines	0.688953	0.0383488	17.97	3.63e-072 ***	
vehicles	-1.52965	0.0418650	-36.54	2.79e-292 ***	
japan	0.222222	0.0275020	8.080	6.46e-016 ***	
us	-0.299507	0.0253000	-11.84	2.48e-032 ***	
Mean dependent var		73.58564	S.D. dependent var		150.9517
Sum squared resid		1530014	S.E. of regression		94.31559
McFadden R-squared		0.675242	Adjusted R-squared		0.674652
Log-likelihood -4950.		-4950.789	Akaike criterion		9919.578
Schwarz criterion 9		9948.365	Hannan-Quinn		9931.249
Overdispersio	on test: Chi-squa	re(1) = 18.6564	[0.0000]		

#### Poisson Regression Model: Overdispersion

Equidispersion condition

Poisson distributed X obeys

 $\mathsf{E}\{X\} = \mathsf{V}\{X\} = \lambda$ 

- In many situations not realistic
- Overdispersion

Remedies: Alternative distributions, e.g., negative Binomial, and alternative estimation procedures, e.g., Quasi-ML, robust standard errors

#### Count Data Models in GRETL

Model > Limited dependent variable > Count data

- Estimates the coefficients β of the specified model using Poisson (Poisson) or the negative binomial (NegBin 1, NegBin 2) distribution
- Performs overdispersion test for Poisson regression

#### Contents

- Limited Dependent Variable Cases
- Binary Choice Models
- Binary Choice Models: Estimation
- Binary Choice Models: Goodness of Fit
- Application to Latent Models
- Multi-response Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- The Tobit Model: Estimation
- The Tobit II Model

#### Tobit Models

Tobit models are regression models where the range of the (continuous) dependent variable is constrained, i.e., censored from below

Examples:

- Expenditures on durable goods as a function of income, age, etc.: a part of the units does not spend any money on durable goods
- Hours of work as a function of age, qualification, etc.
- Expenditures on alcoholic beverages and tobacco
- Holiday expenditures as a function of the number of children
   Tobit models
- Standard Tobit model or Tobit I model; James Tobin (1958) on expenditures on durable goods
- Generalizations: Tobit II to V

# Example: Expenditures on Tobacco

Verbeek's data set TOBACCO: expenditures on tobacco and alcoholic beverages in 2724 Belgian households, Belgian household budget survey of 1995/96

Model:

 $y_i^* = x_i^{\prime}\beta + \varepsilon_i$ 

- y<sub>i</sub><sup>\*</sup>: optimal expenditure (expenditure corresponding to maximal utility) on tobacco in household *i* (latent)
- x<sub>i</sub>: characteristics of the *i*-th household
- ε<sub>i</sub>: unobserved heterogeneity (or measurement error or optimization error)

Actual expenditures y<sub>i</sub>

$$y_i = y_i^* \text{ if } y_i^* > 0$$
  
= 0 if  $y_i^* \le 0$ 

#### The Standard Tobit Model

The latent variable  $y_i^*$  depends upon characteristics  $x_i$ 

$$y_i^* = x_i^{\beta} + \varepsilon_i$$

with error terms (or unobserved heterogeneity)

 $\varepsilon_i \sim \text{NID}(0, \sigma^2)$ , independent of  $x_i$ 

Actual outcome of the observable variable  $y_i$ 

$$y_i = y_i^* \text{ if } y_i^* > 0$$
  
= 0 if  $y_i^* \le 0$ 

- Standard Tobit model or censored regression model
- Censoring: negative values are substituted by zero
- Censoring in general
  - Censoring from below (above): all values left (right) from a lower (an upper) bound are substituted by the lower (upper) bound
- OLS produces inconsistent estimators for β

#### The Standard Tobit Model, cont'd

Standard Tobit model describes

1. the probability  $P\{y_i = 0\}$  as a function of  $x_i$ 

 $\mathsf{P}\{y_i = 0\} = \mathsf{P}\{y_i^* \le 0\} = \mathsf{P}\{\varepsilon_i \le -x_i^*\beta\} = 1 - \Phi(x_i^*\beta/\sigma)$ 

2. the distribution of  $y_i$  given that it is positive, i.e., the truncated normal distribution with expectation

 $\mathsf{E}\{y_i \mid y_i^* > 0\} = x_i'\beta + \mathsf{E}\{\varepsilon_i \mid \varepsilon_i > -x_i'\beta\} = x_i'\beta + \sigma \lambda(x_i'\beta/\sigma)$ with  $\lambda(x_i'\beta/\sigma) = \phi(x_i'\beta/\sigma) / \Phi(x_i'\beta/\sigma) \ge 0$ 

Attention! A single set  $\beta$  of parameters characterizes both expressions

- The effect of a characteristic
  - on the probability of non-zero observation and
  - on the value of the observation

have the same parameter and sign!

### The Standard Tobit Model: Interpretation

From

 $P\{y_i = 0\} = 1 - \Phi(x_i'\beta/\sigma)$  $E\{y_i \mid y_i > 0\} = x_i'\beta + \sigma \lambda(x_i'\beta/\sigma)$ 

follows:

- A positive coefficient  $\beta_k$  means that an increase in the explanatory variable  $x_{ik}$  increases the probability of having a positive  $y_i$
- The marginal effect of  $x_{ik}$  upon E{ $y_i | y_i > 0$ } is different from  $\beta_k$ 
  - Attention: Do not estimate the  $\beta_k$ 's from the positive observations only!
- The marginal effect of  $x_{ik}$  upon E{ $y_i$ } can be shown to be  $\beta_k P{y_i > 0}$ 
  - □ It is close to  $\beta_k$  if P{ $y_i > 0$ } is close to 1, i.e, little censoring
- The marginal effect of  $x_{ik}$  upon E{ $y_i^*$ } is  $\beta_k$  (due to  $y_i^* = x_i^{\beta} + \varepsilon_i$ )
  - These effects are not really of interest
#### Contents

- Limited Dependent Variable Cases
- Binary Choice Models
- Binary Choice Models: Estimation
- Binary Choice Models: Goodness of Fit
- Application to Latent Models
- Multi-response Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- The Tobit Model: Estimation
- The Tobit II Model

### The Standard Tobit Model: Estimation

OLS produces inconsistent estimators for  $\beta$ ; alternatives:

1. ML estimation based on the log-likelihood

 $\log L_1(\beta, \sigma^2) = \ell_1(\beta, \sigma^2) = \sum_{i \in I0} \log P\{y_i = 0\} + \sum_{i \in I1} \log f(y_i)$ 

with appropriate expressions for P{.} and f(.),  $I_0$  the set of censored observations,  $I_1$  the set of uncensored observations

For the correctly specified model: estimates are

- Consistent
- Asymptotically efficient
- Asymptotically normally distributed
- 2. Truncated regression model: ML estimation based on observations with  $y_i > 0$  only (observations with  $y_i = 0$  may be unavailable):

 $\ell_2(\beta, \sigma^2) = \Sigma_{i \in \mathsf{I}1}[ \log f(y_i | y_i > 0)] = \Sigma_{i \in \mathsf{I}1}[ \log f(y_i) - \log \mathsf{P}\{y_i > 0\}]$ 

Estimates based on  $l_1$  are more efficient than those based on  $l_2$ 

## Example: Model for Budget Shares for Tobacco and Alcohol

Verbeek's data set TOBACCO: Belgian household budget survey of 1995/96; expenditures for tobacco and alcoholic beverages

- Budget share  $w_i^*$  for expenditures on alcoholic beverages corresponding to maximal utility:  $w_i^* = x_i^{\prime}\beta + \varepsilon_1$ 
  - $x_i$ : log of total expenditures (LNX) and various characteristics like
  - number of adults in household (NADULTS)
  - □ number of children  $\leq$  2 years old (NKIDS2)
  - age (AGE)

Actual budget share for expenditures on alcohol (SHARE1, W1)

 $w_{i} = w_{i}^{*}$  if  $w_{i}^{*} > 0$ ,

= 0 otherwise

2724 households

# Budget Shares for Tobacco and Alcohol, cont'd

Belgian household budget in 1995/96:

No expenditures

- for alcoholic beverages: 17% of households
- for tobacco: 62% of households

Average budget share

- for alcoholic beverages: 2.15%
- for tobacco: 3.22%

### Model for Budget Share

Budget share  $w_i^*$  for expenditures on alcoholic beverages

$$w_i^* = x_i^{\beta} + \varepsilon_l$$

regressors *x*<sub>i</sub>:

- log of total expenditures (LNX) and
- household characteristics: AGE, NADULTS, NKIDS, NKIDS2
- interactions AGELNX (=LNX\*AGE), NADLNX (=LNX\*NADULTS)

Actual budget share for expenditures on alcohol (SHARE1, W1)

$$w_{i} = w_{i}^{*}$$
 if  $w_{i}^{*} > 0$ ,

= 0 otherwise

Attention! Sufficiently large change of income will create positive *w*\* for any household!

# Model for Budget Share for Alcohol

#### Tobit model, GRETL output

Model 2: Tobit, using observations 1-2724 Dependent variable: SHARE1 (alcohol)							
coefficient		:	std. er	ror	t-ratio	p-value	
const AGE NADULTS NKIDS NKIDS2 LNX AGELNX NADLNX	-0,17041 0,015212 0,02804 -0,00295 -0,00411 0,01343 -0,00094 -0,00218	7 20 18 209 756 88 4668 017	0,044 0,010 0,018 0,000 0,003 0,003 0,003 0,000 0,001	1114 6351 8201 794286 20953 26703 787573 36622	 -3,863 1,430 1,490 -3,717 -1,283 4,113 -1,199 -1,596	0,0001 0,1526 0,1362 0,0002 0,1995 3,90e-0 0,2303 0,1105	*** *** 5 ***
WALLOON	0,00417	202	0,000	980745	4,254	2,10e-0	5 ***
Mean depend Censored ob Log-likelihood Schwarz crite	dent var s d erion -	0,0178 466 4764 -9449,	828 ,153 208	S.D. de sigma Akaike Hannar	pendent v criterion n-Quinn	var 0,0 0,0 -95 -94	21658 24344 508,306 86,944

# Model for Budget Share for Alcohol, cont'd

Truncated regression model, GRETL output

Model 7: Tobit, using observations 1-2724 (n = 2258) Missing or incomplete observations dropped: 466 Dependent variable: W1 (alcohol)

C	oefficient	std. error	t-ratio	p-value
const	0,0433570	0,0458419	0,9458	0,3443
AGE	0,00880553	0,0110819	0,7946	0,4269
NADULTS	-0,0129409	0,0185585	-0,6973	0,4856
NKIDS	-0,00222254	0,000826380	-2,689	0,0072 ***
NKIDS2	-0,00261220	0,00335067	-0,7796	0,4356
LNX	-0,00167130	0,00337817	-0,4947	0,6208
AGELNX	-0,000490197	0,000815571	-0,6010	0,5478
NADLNX	0,000806801	0,00134731	0,5988	0,5493
WALLOON	0,00261490	0,000922432	2,835	0,0046 ***
Mean depend	dent var 0,0215	507 S.D. depe	endent va	r 0,022062
Censored ob	s 0	sigma		0,021450
Log-likelihoo	d 5471,3	304 Akaike cr	iterion	-10922,61
Schwarz crite	erion -10865,3	9 Hannan-0	Quinn	-10901,73

# Models for Budget Share for Alcohol, Comparison

Estimates (coeff.) and standard errors (s.e.) for some coefficients

of the Tobit (2724 observations, 644 censored) and the truncated regression model (2258 uncensored observations)

		constant	NKIDS	LNX	WALL
Tobit model	coeff.	-0,1704	-0,0030	0,0134	0,0042
	s.e.	0,0441	0,0008	0,0033	0,0010
Truncated regression	coeff.	0,0433	-0,0022	-0,0017	0,0026
	s.e.	0,0458	0,0008	0,0034	0,0009

### **Specification Tests**

#### Tests

- for normality of error terms
- for omitted variables
- Tests based on
- generalized residuals

 $\lambda(-x_i'\beta/\sigma)$  if  $y_i = 0$ 

 $e_i/\sigma$  if  $y_i > 0$  (standardized residuals)

with  $\lambda(-x_i'\beta/\sigma) = -\phi(x_i'\beta/\sigma) / \Phi(-x_i'\beta/\sigma)$ , evaluated for estimates of  $\beta$ ,  $\sigma$ 

and "second order" generalized residuals corresponding to the estimation of  $\sigma^2$ 

Test for normality is standard test in GRETL's Tobit procedure: consistency requires normality

#### Contents

- Limited Dependent Variable Cases
- Binary Choice Models
- Binary Choice Models: Estimation
- Binary Choice Models: Goodness of Fit
- Application to Latent Models
- Multi-response Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- The Tobit Model: Estimation
- The Tobit II Model

## An Example: Modeling Wages

Wages: observations available only for the working population

- Model that explains wages as a function of characteristics, e.g., education, the person's age, gender, etc.
- Low value of education increases probability of no wage
  - From a sample of wages the effect of education might be underestimated
  - "Sample selection bias"
- Tobit model: for a positive coefficient of age, an increase of age
  - increases wage
  - increases the probability that the person is working
  - Not always realistic!

Tobacco consumption: Abstention from smoking may be a person's attitude not depending on factors which determine smoking intensity

### Modeling Wages, cont'd

Tobit II model: allows two separate equations:

- Equation for labour force participation of a person
- Equation for the wage of a person

Tobit II model is also called "sample selection model"

### Tobit II Model for Wages

Wage equation describes the wage of person i

 $w_i^* = x_{1i}^{\beta}\beta_1 + \varepsilon_{1i}$ 

with exogenous characteristics (age, education, ...)

Selection equation or labor force participation

 $h_i^* = x_{2i}^{\beta_2} + \varepsilon_{2i}$ 

Observation rule: w<sub>i</sub> actual wage of person i

 $w_{i} = w_{i}^{*}, h_{i} = 1 \text{ if } h_{i}^{*} > 0$ 

 $w_i$  not observed,  $h_i = 0$  if  $h_i^* \le 0$ 

h<sub>i</sub>: indicator for working

Distributional assumption for  $\varepsilon_{1i}$ ,  $\varepsilon_{2i}$ : usually normality with

$$\begin{pmatrix} \boldsymbol{\varepsilon}_{1i} \\ \boldsymbol{\varepsilon}_{2i} \end{pmatrix} \sim N \begin{bmatrix} 0, \begin{pmatrix} \boldsymbol{\sigma}_1^2 & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{12} & \boldsymbol{\sigma}_2^2 \end{bmatrix} \end{bmatrix}$$

# Model for Wages: Selection Equation

Selection equation  $h_i^* = x_{2i}^{\ \beta_2} + \varepsilon_{2i}^{\ \beta_2}$  probit model for binary choice; standardization ( $\sigma_2^2 = 1$ )

- Characteristics  $x_{1i}$  and  $x_{2i}$  may be different; however,
  - The selection may depend upon  $w_i^*$ :  $x_{2i}$  is expected to include  $x_{1i}$
  - $x_{2i}$  should contain variables not included in  $x_{1i}$
  - Sign and value of coefficients of the same variables in x<sub>1i</sub> and x<sub>2i</sub> are not the same
- Special cases
  - If  $\sigma_{12} = 0$ , sample selection is exogenous
  - Tobit II model coincides with Tobit I model if  $x_{1i}^{\,\,\prime}\beta_1 = x_{2i}^{\,\,\prime}\beta_2$  and  $\varepsilon_{1i} = \varepsilon_{2i}^{\,\,\prime}$

# Model for Wages: Wage Equation

Expected value of  $w_i$ , given sample selection:

 $\mathsf{E}\{w_{i} \mid h_{i} = 1\} = x_{1i}'\beta_{1} + \sigma_{12}\lambda(x_{2i}'\beta_{2})$ 

with the inverse Mill's ratio or Heckman's lambda

 $\lambda(x_{2\mathrm{i}}`\beta_2) = \phi(x_{2\mathrm{i}}`\beta_2) \,/\, \Phi(x_{2\mathrm{i}}`\beta_2)$ 

- Heckman's lambda
  - Positive and decreasing in its argument
  - The smaller the probability that a person is working, the larger the value of the correction term  $\lambda$
- Expected value of  $w_i$  only equals  $x_{1i}$ ' $\beta_1$  if  $\sigma_{12} = 0$ : no sample selection error, consistent OLS estimates of the wage equation

### Tobit II Model: Log-likelihood Function

Log-likelihood

 $\begin{aligned} \ell_{3}(\beta_{1},\beta_{2},\sigma_{1}^{2},\sigma_{12}) &= \sum_{i \in I0} \log \mathsf{P}\{h_{i}=0\} + \sum_{i \in I1} [\log f(y_{i}|h_{i}=1) + \log \mathsf{P}\{h_{i}=1\}] \\ &= \sum_{i \in I0} \log \mathsf{P}\{h_{i}=0\} + \sum_{i \in I1} [\log f(y_{i}) + \log \mathsf{P}\{h_{i}=1|y_{i}\}] \end{aligned}$ 

with

$$P\{h_{i}=0\} = 1 - \Phi(x_{2i}'\beta_{2})$$

$$f(y_{i}) = \frac{1}{\sqrt{2\pi\sigma_{1}^{2}}} \exp\left\{-\frac{1}{2\sigma_{1}^{2}}(y_{i} - x_{1i}'\beta_{1})^{2}\right\}$$

$$P\{h_{i}=1|y_{i}\} = \Phi\left(\frac{x_{2i}'\beta_{2} + (\sigma_{12}/\sigma_{1}^{2})(y_{i} - x_{1i}'\beta_{1})}{\sqrt{1 - \sigma_{12}^{2}/\sigma_{1}^{2}}}\right)$$

and using  $f(y_i|h_i = 1) P\{h_i = 1\} = P\{h_i = 1|y_i\} f(y_i)$ 

#### Tobit II Model: Estimation

Maximum likelihood estimation, based on the log-likelihood

 $\ell_{3}(\beta_{1},\beta_{2},\sigma_{1}^{2},\sigma_{12}) = \sum_{i \in I0} \log P\{h_{i}=0\} + \sum_{i \in I1} [\log f(y_{i}|h_{i}=1) + \log P\{h_{i}=1\}]$ 

- Two step approach (Heckman, 1979)
  - 1. Estimate the coefficients  $\beta_2$  of the selection equation by standard probit maximum likelihood:  $b_2$
  - 2. Compute estimates of Heckman's lambda:

 $\lambda_i = \lambda(x_{2i}, b_2) = \phi(x_{2i}, b_2) / \Phi(x_{2i}, b_2)$  for i = 1, ..., N

3. Estimate the coefficients  $\beta_1$  and  $\sigma_{12}$  using OLS

 $w_i = x_{1i}\beta_1 + \sigma_{12}\lambda_i + \eta_i$ 

 GRETL: procedure "Heckit" allows both the ML and the two step estimation

### Tobit II Model for Budget Share for Alcohol

Heckit ML estimation, GRETL output	Model 7: ML Heckit, using observations 1-2724 Dependent variable: SHARE1 Selection variable: D1 coefficient std. error t-ratio p-value
D1: dummy, 1 if SHARE1 > 0	const0,04441780,04924400,90200,3671AGE0,008743700,01102720,79290,4278NADULTS-0,01308980,0165677-0,79010,4295NKIDS-0,002217650,000585669-3,7870,0002***NKIDS2-0,002601860,00228812-1,1370,2555LNX-0,001745570,00357283-0,48860,6251AGELNX-0,0004858660,000807854-0,60140,5476NADLNX0,0008178260,001195740,68390,4940WALLOON0,002605570,0009585042,7180,0066***lambda-0,000137730,00291516-0,047250,9623
	Mean dependent var         0,021507         S.D. dependent var         0,022062           sigma         0,021451         rho         -0,006431           Log-likelihood         4316,615         Akaike criterion         -8613,231           Schwarz criterion         -8556,008         Hannan-Quinn         -8592,349

## Tobit II Model for Budget Share for Alcohol, cont'd

Heckit ML					
estimation, GRETL output					
	Selection equation coefficient std. error t-ratio p-value				
	const	-16,2535	2,58561	-6,286	3,25e-10 ***
	AGE	0,753353	0,653820	1,152	0,2492
	NADULTS	2,13037	1,03368	2,061	0,0393 **
	NKIDS	-0,0936353	0,0376590	-2,486	0,0129 **
	NKIDS2	-0,188864	0,141231	-1,337	0,1811
	LNX	1,25834	0,192074	6,551	5,70e-11 ***
	AGELNX	-0,0510698	0,0486730	-1,049	0,2941
	NADLNX	-0,160399	0,0748929	-2,142	0,0322 **
	WHITECOL	-0,0352022 L 0,0801599 I 0,201073	0,0983073 0,0852980 0,0628750	-0,3581 0,9398 3,198	0,7203 0,3473 0,0014 ***

### Models for Budget Share for Tabacco

Estimates and standard errors for some coefficients of the standard Tobit, the truncated regression and the Tobit II model

		const.	NKIDS	LNX	WALL
Tobit model	coeff.	-0,1704	-0,0030	0,0134	0,0042
	s.e.	0,0441	0,0008	0,0033	0,0010
Truncated regression	coeff.	0,0433	-0,0022	-0,0017	0,0026
	s.e.	0,0458	0,0008	0,0034	0,0009
Tobit II	coeff.	0,0444	-0,0022	-0,0017	0,0026
model	s.e.	0,0492	0,0006	0,0036	0,0010
Tobit II	coeff.	-16,2535	-0,0936	1,2583	0,2011
selection	s.e.	2,5856	0,0377	0,1921	0,0629

### Test for Sampling Selection Bias

Error terms of the Tobit II model with  $\sigma_{12} \neq 0$ : standard errors and test may result in misleading inferences

- Test of H<sub>0</sub>:  $\sigma_{12} = 0$  in the second step of Heckit, i.e., fitting the regression  $w_i = x_{1i}\beta_1 + \sigma_{12}\lambda_i + \eta_i$
- GRETL: *t*-test on the coefficient for Heckman's lambda
- GRETL: Heckit-output shows rho, estimate for  $\rho_{12}$  from  $\sigma_{12} = \rho_{12}\sigma_1$
- Test results are sensitive to exclusion restrictions on  $x_{1i}$

#### Tobit Models in GRETL

Model > Limited dependent variable > Tobit

- Estimates the Tobit model; censored dependent variable
  Model > Limited dependent variable > Heckit
- Estimates in addition the selection equation (Tobit II), optionally by ML- and by two-step estimation

#### Your Homework

- 1. Verbeek's data set TOBACCO contains expenditures on tobacco in 2724 Belgian households, taken from the household budget survey of 1995/96, as well as other characteristics of the households; for the expenditures on tobacco, the dummy D2=1 if the budget share for tobacco (SHARE2) differs from 0, and D2=0 otherwise.
  - a. Model the budget share for tobacco, using (i) a Tobit model, (ii) a truncated regression, and (iii) a Tobit II model; using the household characteristics LNX, AGE, NKIDS, the interaction LNX\*AGE, and the dummy FLANDERS; in addition NADULTS and AGE2 for the selection equation.
  - b. Compare the effects of the regressors in the three models, based on coefficients and *t*-statistics.
  - c. Discuss the effect of the variable FLANDERS.

#### Your Homework, cont'd

- 2. People buy for  $y_i^*$  assets of an investment fund, with  $y_i^* = x_i^{\beta} + \varepsilon_i$ ,  $\varepsilon_i \sim N(0, \sigma^2)$ ;  $x_i = (1, income)$ '. The dummy  $d_i = 1$  if  $y_i^* > 0$  and  $d_i = 0$  otherwise.
  - a. Derive the probability for  $d_i = 1$  as function of  $x_i$ .
  - b. Derive the log-likelihood function of the probit model for  $d_i$ , i = 1,...,N.
  - c. Derive the first-order condition related to the maximization of the loglikelihood function for the logit model; cf. equation (A) on slide 20.
  - d. Derive the ML estimator of the probability for  $d_i = 1$  as function of  $x_i$  of the logit model.