Econometrics 2 - Lecture 2

Models with Limited Dependent Variables

Contents

Cases of Limited Dependent Variable

Typical situations: function of explanatory variables are of interest to explain

- Dichotomous 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
  - Proportion of car owning households:19/32 = 0.59
- But: this probability will differ for rich and poor!
- The sample data has 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 prediction of car ownership take the income of a household into account?

Notation: From N households

- dummy y_i for car ownership: y_i =1: household i has car
- income x_{i2}

For predicting y_i – or of P{y_i =1} – , a model is needed that takes the income into account

Modeling Car Ownership

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

1. Linear regression $x_i'\beta + \varepsilon_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i$ for describing y

- With $E\{\varepsilon_i|x_i\} = 0$, the model $y_i = x_i'\beta + \varepsilon_i$ gives

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

  due to $E\{y_i|x_i\} = 1*P\{y_i = 1|x_i\} + 0*P\{y_i = 0|x_i\} = P\{y_i = 1|x_i\}$
- Model $y_i = x_i'\beta + \varepsilon_i$: $x_i'\beta$ can be interpreted as $P\{y_i = 1|x_i\}$!
- Problems:
  - $x_i'\beta$ not necessarily in [0,1]
  - Error terms: for a given x_i
    - ε_i has only two values, viz. $1 - x_i'\beta$ and $x_i'\beta$
    - $V\{\varepsilon_i|x_i\} = x_i'\beta(1 - x_i'\beta)$, heteroskedastic, dependent upon β
- Model for y actually is specifying the probability that $y=1$ as a function of x

Modeling Car Ownership, cont'd

2. Use of a function G(x_i,β) with values in the interval [0,1]

   P{y_i =1|x_i} = E{y_i|x_i} = G(x_i,β)

- The probability that y_i =1, i.e., the household owns a car, depends on the income (and other characteristics, e.g., family size)
- Use for G(x_i,β) the standard logistic distribution function

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

  F(*z*) fulfills $\lim_{z\to -\infty} F(z) = 0$, $\lim_{z\to \infty} F(z) = 1$
- Interpretation:
  - From P{y_i =1|x_i} = p_i = exp{x_i'β}/(1+exp{x_i'β}) follows

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

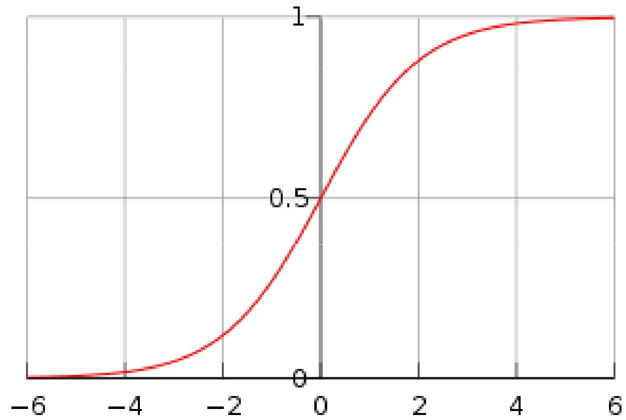
  - An increase of x_{i2} by 1 results in a relative change of the odds $p_i/(1- p_i)$ by β_2 or by $100\beta_2$%; cf. the notion semi-elasticity

Car Ownership and Income, cont'd

E.g., P{y_i =1|x_i} = 1/(1+exp(-z_i)) with z = -0.5+1.1*x, the income 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 3000, z = 3.1 and the probability of owning a car is 1/(1+exp(-3.1)) = 0.94

The standard logistic distribution function, with z on the horizontal and F(z) on the vertical axis



Odds

The odds in favor of an event are the ratio of a pair of integers, the first (the second) representing the relative likelihood that the event will happen (will not happen)

- If p is the probability in favor of the event, the probability against the event therefore being 1-p, the odds of the event are the quotient $\frac{p}{1-p}$
- Example: the odds that a randomly chosen day of the week is a Sunday are 1:6 (say “one to six”) because p = P{Sunday} = 1/7, $p/(1-p) = (1/7)/(6/7) = 1/6$
- In bookmakers language: odds are not in favor but against
  - The bookmaker would say: “The odds that a randomly chosen day of the week is a Sunday are 6:1”
- The logarithm of the odds is the logit of the probability

Contents

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 β, such as

$$E\{y_i|x_i\} = P\{y_i = 1|x_i\} = G(x_i,\beta)$$

Typical functions G(x_i,β): distribution functions (cdf's) F(x_i'β)

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

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

- 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)

$$\begin{aligned} F(z) &= 0,\ z < 0 \\ &= z,\ 0 \le z \le 1 \\ &= 1,\ z > 1 \end{aligned}$$

Linear Probability Model (LPM)

Assumes that

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

but sets

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

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

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

Probit Model: Standardization

$E\{y_i|x_i\} = P\{y_i = 1|x_i\} = G(x_i,\beta)$: assume $G(.)$ to be $N(0, \sigma^2)$

$$P\left\{y_i = 1 \middle| x_i\right\} = \Phi\left(\frac{x_i'\beta}{\sigma}\right)$$

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

Probit vs Logit Model

- Differences between the probit and the logit model:
  - Shape of distribution is slightly different, particularly in the tails.
  - Scaling of the distribution is different: The implicit variance for ε_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 β not directly comparable across the two models, while the signs are typically the same
  - The estimates in the logit model are roughly a factor $\pi/\sqrt{3} \approx 1.81$ larger than those in the probit model

Interpretation of Coefficients

For assessing the effect of changing x_k the

- Coefficient β_k

is of interest, but also related characteristics such as

- Sign
- Slope, i.e., the "average" marginal effect $\partial F(x_i'\beta)/\partial x_{ik}$

Binary Choice Models: Marginal Effects

Linear regression models: β_k is the marginal effect of a change in x_k

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

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

with density function f(.)

- The effect of changing the regressor x_k depends upon $x_i'\beta$, the shape of F, and β_k
- The marginal effect of changing x_k
  - Probit model: $\phi(x_i'\beta)\,\beta_k$, with standard normal density function ϕ
  - Logit model: $L(x_i'\beta)[1 - L(x_i'\beta)]\,\beta_k$
  - Linear probability model

$$\frac{\partial x_i'\beta}{\partial x_{ik}} = \beta_k, \ \ if \ \ x_i'\beta \in [0,1]$$

Binary Choice Models: Slopes

Interpretation of the effect of a change in x_k

- "Slope", i.e., the gradient of E{$y_i|x_i$} at the sample means of the regressors

$$slope_k(\bar{x}) = \left.\frac{\partial F(x_i'\beta)}{\partial x_k}\right|_{\bar{x}}$$

- 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 β_k is the relative change of the odds 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
- Multiresponse Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- The Tobit II Model

Binary Choice Models: Estimation

Typically, binary choice models are estimated by maximum likelihood

Likelihood function

$$L(\beta) = \Pi_{i=1}^{N} P\{y_i = 1|x_i;\beta\}^{y_i} \, P\{y_i = 0|x_i;\beta\}^{1-y_i}$$
$$= \Pi_i \, F(x_i'\beta)^{y_i} \, (1 - F(x_i'\beta))^{1-y_i}$$

- Maximization via the log-likelihood function

$$\ell(\beta) = \log L(\beta) = \Sigma_i \, y_i \log F(x_i'\beta) + \Sigma_i \, (1-y_i) \log (1-F(x_i'\beta))$$

- First-order conditions of the maximization problem

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_i \left[\frac{y_i - F(x_i'\beta)}{F(x_i'\beta)(1-F(x_i'\beta))} f(x_i'\beta) \right] x_i = \sum_i e_i x_i = 0$$

- e_i: generalized residuals

Generalized Residuals

The first-order conditions allow to define the generalized residuals

From

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_i \left[\frac{y_i - F(x_i'\beta)}{F(x_i'\beta)(1-F(x_i'\beta))} f(x_i'\beta) \right] x_i = \sum_i e_i x_i = 0$$

- follows that the generalized residuals e_i can assume two values:
  - $e_i = f(x_i'b)/F(x_i'b)$ if $y_i = 1$
  - $e_i = - f(x_i'b)/(1-F(x_i'b))$ if $y_i = 0$

  b are the estimates of β
- 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

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_i \left[y_i - \frac{\exp(x_i'\beta)}{1-\exp(x_i'\beta)} \right] x_i = 0$$

gives

$$\hat{p}_i = \frac{\exp(x_i'b)}{1+\exp(x_i'b)}$$

- From $\Sigma_i \hat{p}_i x_i = \Sigma_i y_i x_i$ follows – given one regressor is an intercept –:
  - The predicted frequency $\Sigma_i \hat{p}_i$ equals the observed frequency $\Sigma_i y_i$
- Similar results for the probit model, due to similarity of logit and probit functions

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 between errors and regressors
- No omitted regressors

Contents

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^2
  - *McFadden* R^2
- Index based on proportion of correctly predicted observations
  - Hit rate

McFadden R^2

Based on log-likelihood function

- $\ell(b) = \ell_1$: maximum log-likelihood of the model to be assessed
- ℓ_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 $\ell_1 - \ell_0$, the more contribute the regressors
  - $\ell_1 = \ell_0$, if all slope coefficients are zero
  - $\ell_1 = 0$, if y_i is exactly predicted for all i
- *Pseudo*-R^2: a number in [0,1), defined by

$$pseudo - R^2 = 1 - \frac{1}{1 + 2(\ell_1 - \ell_0)/N}$$

- *McFadden* R^2: a number in [0,1], defined by

$$McFadden R^2 = 1 - \ell_1 / \ell_0$$

- Both are 0 if $\ell_1 = \ell_0$
- *McFadden* R^2 attains the upper limit if $\ell_1 = 0$

Hit Rate

Comparison of correct and incorrect predictions

- Predicted outcome

  $\hat{y}_i = 1$ if $x_i'b > 0$

  $= 0$ if $x_i'b \leq 0$
- 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

    $\hat{y}_i = 1$ if $\hat{p}_i = N_1/N > 0.5$, $\hat{y}_i = 0$ if $\hat{p}_i \leq 0.5$
  - $R_p^2 = 1 – wr_1/wr_0$

    with $wr_0 = 1 - \hat{p}_i$ if $\hat{p}_i > 0.5$, $wr_0 = \hat{p}_i$ if $\hat{p}_i \leq 0.5$ in order to avoid $R_p^2 < 0$

	$\hat{y} = 0$	$\hat{y} = 1$	Σ
$y = 0$	n_{00}	n_{01}	N_0
$y = 1$	n_{10}	n_{11}	N_1
Σ	n_0	n_1	N

Example: Effect of Teaching Method

Study by Spector & Mazzeo (1980); see Greene (2003), Chpt.21

Personalized System of Instruction: new teaching method in economics; has it 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 pretest, entering knowledge
- 32 observations

Effect of Teaching Method, cont'd

Logit model for GRADE, GRETL output

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

	Coefficient	*Std. Error*	*z-stat*	*Slope**
const	-13.0213	4.93132	-2.6405	
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

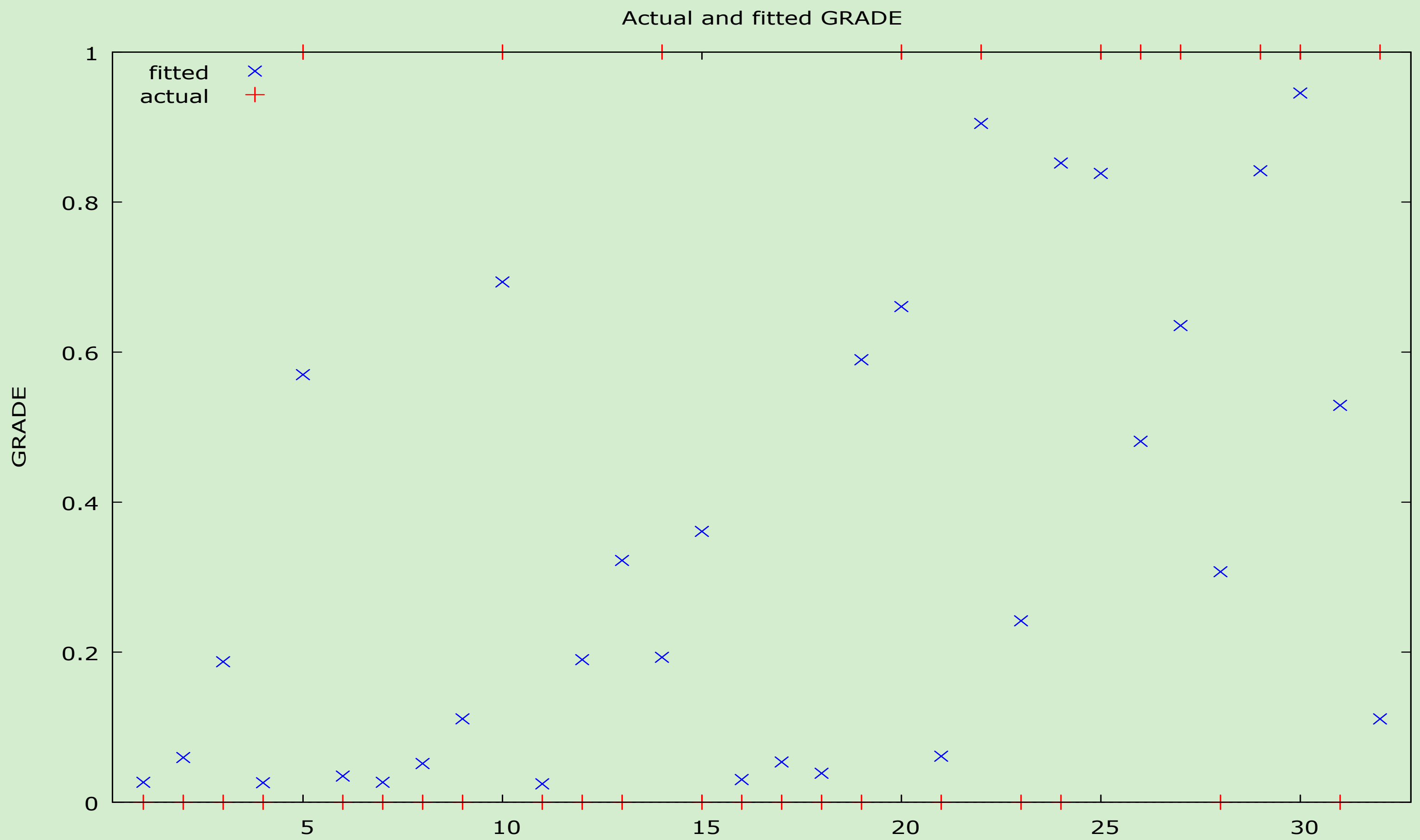
Mean dependent var	0.343750	S.D. dependent var	0.188902
McFadden R-squared	0.374038	Adjusted R-squared	0.179786
Log-likelihood	-12.88963	Akaike criterion	33.77927
Schwarz criterion	39.64221	Hannan-Quinn	35.72267

*Number of cases 'correctly predicted' = 26 (81.3%)
f(beta'x) at mean of independent vars = 0.189
Likelihood ratio test: Chi-square(3) = 15.4042 [0.0015]

		Predicted	
		0	1
Actual	0	18	3
	1	3	8

Effect of Teaching Method, cont'd

Logit model for GRADE, actual and fitted values of 32 observations



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	s.e.	coeff	s.e.	coeff	s.e.
const	-1.498	0.524	-13.02	4.931	-7.452	2.542
GPA	0.464	0.162	2.826	1.263	1.626	0.694
TUCE	0.010	0.019	0.095	0.142	0.052	0.084
PSI	0.379	0.139	2.379	1.065	1.426	0.595

- Coefficients of logit model: due to larger variance, larger by factor $\sqrt{(\pi^2/3)}=1.81$ than that of the probit model

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: $\ell_0 = -12.89$, $wr_1 = 6/32$

Goodness of fit measures

- $R_p^2 = 1 – wr_1/wr_0 = 1 – 6/11 = 0.45$
- McFadden $R^2 = 1 – (-12.89)/(-20.59) = 0.374$

Example: Utility of Car Owning

Latent variable y_i^*: utility difference between owning and not owning a car; unobservable (latent)

- Decision on owning a car
  - $y_i^* > 0$: in favor of car owning
  - $y_i^* \leq 0$: against car owning
- y_i^* depends upon observed characteristics (like income) and unobserved characteristics ε_i

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

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

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

last step requires a symmetric distribution function F(.)

Latent variable model: based on a latent variable that represents underlying behavior

Latent Variable Model

Model for the latent variable y_i^*

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

y_i^*: not necessarily a utility difference

- ε_i's are independent of x_i's
- ε_i has standardized distribution
  - Probit model if ε_i has standard normal distribution
  - Logit model if ε_i has standard logistic distribution
- Observations
  - $y_i = 1$ if $y_i^* > 0$
  - $y_i = 0$ if $y_i^* \leq 0$
- ML estimation

Binary Choice Models in GRETL

Model > Nonlinear Models > Logit > Binary

- Estimates the specified model using error terms with standard logistic distribution

Model > Nonlinear Models > Probit > Binary

- Estimates the specified model using error terms with standard normal distribution

Contents

Multiresponse Models

Model 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.
- Multiresponse models describe the probability of each of these outcomes, as a function of variables like
  - person-specific characteristics
  - alternative-specific characteristics
- Types of multiresponse 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 explanatory variables x_i

  $$y_i = j \text{ if } \gamma_{j-1} < y_i^* \leq \gamma_j \text{ for } j = 0,\ldots,M$$

- boundaries γ_j, $j = 0,\ldots,M$, with $\gamma_0 = -\infty, \ldots, \gamma_M = \infty$
- ε_i's are independent of x_i's
- ε_i typically follow the
  - standard normal distribution: ordered probit model
  - standard logistic distribution: ordered logit model

Example: Willingness to Work

„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 the desired labor 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

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

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

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

- ε_i‘s with distribution function F(.)
- y_i^* stands for “willingness to work” or “desired hours of work”

Willingness to Work, cont'd

In terms of observed quantities:

$$P\{y_i = 1 \mid x_i\} = P\{y_i^* \leq 0 \mid x_i\} = F(- x_i'\beta)$$

$$P\{y_i = 3 \mid x_i\} = P\{y_i^* > \gamma \mid x_i\} = 1 - F(\gamma - x_i'\beta)$$

$$P\{y_i = 2 \mid x_i\} = F(\gamma - x_i'\beta) – F(- x_i'\beta)$$

- Unknown parameters: γ and β
- Standardization: wrt location ($\gamma_1 = 0$) and scale ($V\{\varepsilon_i\} = 1$)
- ML estimation

Interpretation of parameters β

- Wrt y_i^*: willingness to work increases with larger x_k for positive β_k
- Wrt probabilities $P\{y_i = j \mid x_i\}$, e.g., $P\{y_i = 3 \mid x_i\}$ increases and $P\{y_i = 1 \mid x_i\}$ decreases with larger x_k for positive β_k

Example: Credit Rating

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

Rating models:

1. Ordered logit model for assignment of categories “1“, …,“7“ (highest)
2. 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	1	7
investment grade	0.472	0	0	1
book leverage	0.293	0.264	0.000	0.999
working capital/total assets	0.140	0.123	−0.412	0.748
retained earnings/total assets	0.157	0.180	−0.996	0.980
earnings before interest and taxes/t.a.	0.094	0.090	−0.384	0.652
log sales	7.996	7.884	1.100	12.701

Book leverage: ratio of debts to assets

Credit Rating, cont'd

Verbeek, Table 7.5.

Table 7.5 Estimation results binary and ordered logit, MLE

	Binary logit Estimate	Binary logit Standard error		Ordered logit Estimate	Ordered logit Standard error
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 *sales*	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
			γ_1	−0.369	0.633
			γ_2	4.881	0.521
			γ_3	7.626	0.551
			γ_4	9.885	0.592
			γ_5	12.883	0.673
			γ_6	14.783	0.784
loglikelihood	−341.08			−965.31	
McFadden R^2	0.465			0.309	
LR test (χ^2_5)	591.8 ($p = 0.000$)			862.9 ($p = 0.000$)	

Ordered Response Model: Estimation

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

- Loglikelihood 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
- Multiresponse Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- The Tobit II Model

Multinomial Models

Choice between M alternatives without natural order

Observed alternative for sample unit i: y_i

"Random utility" framework: Individuals

- attach utility levels U_{ij} to each of the alternatives, $j = 1,\ldots, M$
- choose the alternative with the highest utility level

Utility levels U_{ij}, $j = 1,\ldots, M$, as a function of characteristics x_{ij}

$$U_{ij} = x_{ij}'\beta + \varepsilon_{ij}$$

- error terms ε_{ij} follow the Type I extreme value distribution:

$$P\{y_i = j\} = \frac{\exp\{x_{ij}'\beta\}}{\exp\{x_{i1}'\beta\} + \ldots + \exp\{x_{iM}'\beta\}}$$

for $j = 1, \ldots, M$

- and $\Sigma_j P\{y_i = j\} = 1$

Variants of the Logit Model

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

Conditional logit model: for $j = 1, \ldots, M$

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

- Alternative-specific characteristics x_{ij}
- E.g., mode of transportation is affected by travel costs, travel duration, etc.

Multinomial logit model: for $j = 1, \ldots, M$

$$P\{y_i = j\} = \frac{\exp\{x_i'\beta_j\}}{1 + \exp\{x_i'\beta_j\} + \ldots + \exp\{x_i'\beta_j\}}$$

- Person-specific characteristics x_i
- 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 the mixed logit model: combines
  - Alternative-specific characteristics and
  - Person-specific characteristics

Independence of Errors

Independence of the error terms ε_i implies independent utility levels of alternatives

- A restrictive assumption
- Examples: 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 the number of alternatives: “independence of irrelevant alternatives” (IIA)

Multiresponse Models in GRETL

Model > Nonlinear Models > Logit > Ordered

- Estimates the specified model using error terms with standard logistic distribution, assuming ordered alternatives for responses

Model > Nonlinear Models > Logit > Multinomial

- Estimates the specified model using error terms with standard logistic distribution, assuming alternatives without order

Model > Nonlinear Models > Probit > Ordered / Multinomial

- Estimates the specified model using error terms with standard normal distribution, assuming alternatives with or without order

Contents

- Limited Dependent Variable Cases
- Binary Choice Models
- Binary Choice Models: Estimation
- Binary Choice Models: Goodness of Fit
- Multiresponse Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- 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 misspellings in an email
- Number of applications of a firm for a patent, as a function of
  - Firm size
  - R&D expenditures
  - Industrial sector
  - Country, etc.

  See Verbeek‘s data set PATENT

Poisson Regression Model

Observed variable for sample unit *i*:

y_i: number of possible outcomes 0, 1, …, *y*, …

Aim: to explain $E\{y_i \mid x_i\}$, based on characteristics x_i

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

Poisson regression model

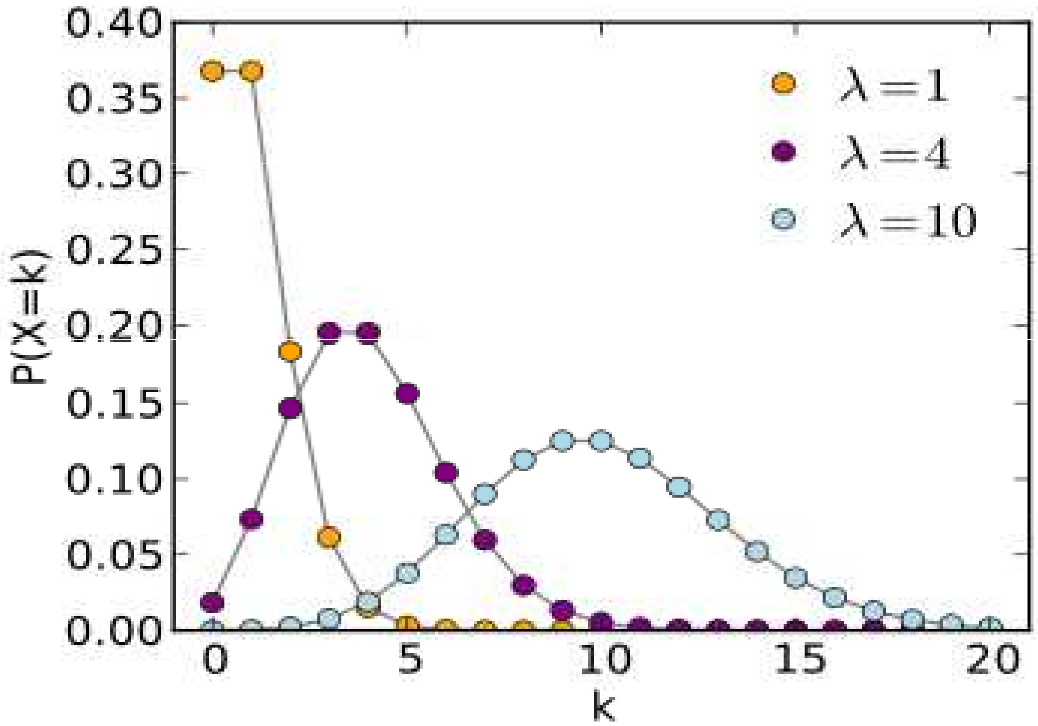
$$P\{y_i = y \mid x_i\} = \frac{\lambda_i^y}{y!}\exp\{\lambda_i\},\ y = 0,1,...$$

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

$y! = 1 \times 2 \times \ldots \times y$, $0! = 1$

Poisson Distribution

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



Poisson Regression Model: The Practice

Unknown parameters: coefficients β

Fitting the model to data: ML estimators are

- Consistent
- Asymptotically efficient
- Asymptotically normally distributed

Equidispersion condition

- Poisson distributed X obeys

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

- In many situations not realistic
- Overdispersion

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

Contents

Tobit Models

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

Examples:

- Expenditures on durable goods as a function of income, age, etc.: a part of units does not spend any money on durable goods
- Hours of work as a function of qualification, age, etc.
- Expenditures on alcoholic beverages and tabacco

Tobit models

- Standard Tobit model or Tobit I model; Tobin (1958) on expenditures on durable goods
- Generalizations: Tobit II to V

Example: Expenditures on Tobacco

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

Model:

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

with optimal expenditures y_i^* on tobacco in household *i*, characteristics x_i of the household, unobserved heterogeneity ε_i (or measurement error or optimization error)

Actual expenditures y_i

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

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)

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

Actual outcome of the observable variable y_i

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

- Standard Tobit model or censored regression model
- Censoring: all 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

   $$P\{y_i = 0\} = P\{\varepsilon_i \leq - x_i'\beta \} = 1 - \Phi(x_i'\beta/\sigma)$$

2. The distribution of y_i given that it is a positive or (by $y_i > 0$) truncated normal distribution

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

   with $\lambda(x_i'\beta/\sigma) = \phi(x_i'\beta/\sigma) \,/\, \Phi(x_i'\beta/\sigma) \geq 0$

Attention: a single set β 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 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 means that an increase in the explanatory variable increases the probability of having a positive y_i
- The marginal effect of x_{ik} upon $E\{y_i \mid y_i > 0\}$ is different from β_k
- The marginal effect of x_{ik} upon $E\{y_i\}$ is $\beta_k P\{y_i > 0\}$
  - It is close to β_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 β_k

The Standard Tobit Model: Estimation

OLS produces inconsistent estimators for β

ML estimation based on the log-likelihood

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

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

ML estimation based on observations with $y_i > 0$ only, i.e., on the truncated regression model:

$$\ell_2(\beta, \sigma^2) = \Sigma_{i\epsilon I1}[\log f(y_i) - \log P\{y_i > 0\}]$$

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

Example: Model for Budget Share for Tobacco

Verbeek's data set TOBACCO: Belgian household budget survey of 1995/96

Budget share w_i^* for expenditures on tobacco corresponding to maximal utility: $w_i^* = x_i'\beta + \varepsilon_I$

x_i: log of total expenditures and various characteristics like

- number of children ≤ 2 years old
- number of adults in household
- age

Actual budget share for expenditures on tobacco

$$w_i = w_i^* \text{ if } w_i^* > 0,$$
$$= 0 \text{ otherwise}$$

- 2724 households

Model for Budget Share for Tobacco

Tobit model, GRETL output

Model 2: Tobit, using observations 1-2724
Dependent variable: SHARE1 (Tobacco)

	coefficient	std. error	t-ratio	p-value	
const	-0,170417	0,0441114	-3,863	0,0001	***
AGE	0,0152120	0,0106351	1,430	0,1526	
NADULTS	0,0280418	0,0188201	1,490	0,1362	
NKIDS	-0,00295209	0,000794286	-3,717	0,0002	***
NKIDS2	-0,00411756	0,00320953	-1,283	0,1995	
LNX	0,0134388	0,00326703	4,113	3,90e-05	***
AGELNX	-0,000944668	0,000787573	-1,199	0,2303	
NADLNX	-0,00218017	0,00136622	-1,596	0,1105	
WALLOON	0,00417202	0,000980745	4,254	2,10e-05	***

Mean dependent var	0,017828	S.D. dependent var	0,021658
Censored obs	466	sigma	0,024344
Log-likelihood	4764,153	Akaike criterion	-9508,306
Schwarz criterion	-9449,208	Hannan-Quinn	-9486,944

Model for Budget Share for Tobacco, 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 (Tobacco)

	coefficient	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 dependent var	0,021507	S.D. dependent var	0,022062
Censored obs	0	sigma	0,021450
Log-likelihood	5471,304	Akaike criterion	-10922,61
Schwarz criterion	-10865,39	Hannan-Quinn	-10901,73

Two Models for Budget Share for Tobacco, Comparison

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

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

Specification Tests

Various tests based on

- generalized residuals

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

  e_i/σ 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 β, σ
- and "second order" generalized residuals corresponding the estimation of σ^2

Tests

- for normality
- for heteroskedasticity
- for omitted variables

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
- Multiresponse Models
- Multinomial Models
- Count Data Models
- The Tobit Model
- The Tobit II Model

An Example: Wage Equation

Wage observations: available only for the working population

Model that explains wages as a function of characteristics, e.g., the person‘s age

- Tobin model: for a positive coefficient of age, an increase of age
  - increases wage
  - increases the probability that the person is working
  - Not always realistic!
- Tobin II model: allows two separate equations
  - for labor force participation and
  - for the wage of a person
- Tobin II model is also called “sample selection model”

Wage Model: Tobit II

- Wage equation describes the wage of person *i*

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

  with exogenous characteristics (age, education, …)
- Labor force participation or selection equation

  $h_i^* = x_{2i}'\beta_2 + \varepsilon_{2i}$
- Observation rule: w_i actual wage of person *i*

  $w_i = w_i^*$, $h_i = 1$ if $h_i^* > 0$

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

  h_i: indicator for working
- Distributional assumption for ε_{1i}, ε_{2i}

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim N\left[0, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right]$$

Wage Model: Selection Equation

- Selection equation: a binary choice model; probit model needs standardization ($\sigma_2^2 = 1$)
- Special cases
  - If $\sigma_{12} = 0$, sample selection is exogenous
  - If $x_{1i}'\beta_1 = x_{2i}'\beta_2$ and $\varepsilon_{1i} = \varepsilon_{2i}$, the Tobit II model coincides with the Tobit I model
- Characteristics x_{1i} and x_{2i} may be different; however,
  - If the selection depends upon w_i^*: x_{2i} is expected to include x_{1i}
  - Because the model describes the joint distribution of w_i and h_i given one set of conditioning variables: x_{2i} is expected to include x_{1i}
  - Sign and value of coefficients of the same variables in x_{1i} and x_{2i} can be different

Wage Model: Wage Equation

Expected value of w_i, given sample selection:

$$E\{w_i \mid h_i = 1\} = x_{1i}'\beta_1 + \sigma_{12}\,\lambda(x_{2i}'\beta_2)$$

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

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

- 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 λ
- Expected value of w_i only equals $x_{1i}'\beta_1$ if $\sigma_{12} = 0$: "no sample selection"

Tobit II Model: Log-likelihood Function

Log-likelihood

$$\ell_3(\beta,\sigma_1^2,\sigma_{12}) = \Sigma_{i\in I0} \log P\{h_i=0\} + \Sigma_{i\in I1} [\log f(y_i|h_i=1) + \log P\{h_i=1\}]$$

$$= \Sigma_{i\in I0} \log P\{h_i=0\} + \Sigma_{i\in I1} [\log f(y_i) + \log P\{h_i=1|y_i\}]$$

with

$$P\{h_i=0\} = 1 - \Phi(x_{2i}'\beta_2)$$

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{1}{2\sigma_1^2}(y_i - x_{1i}'\beta_1)^2\right\}$$

$$P\{h_i = 1 | y_i\} = \Phi\left(\frac{x_{2i}'\beta_2 + (\sigma_{12}/\sigma_1^2)(y_i - x_{1i}'\beta_1)}{\sqrt{1-\sigma_{12}^2/\sigma_1^2}}\right)$$

Tobit II Model: Estimation

- Maximum likelihood estimation, based on the log-likelihood

  $$\ell_3(\beta,\sigma_1^2,\sigma_{12}) = \Sigma_{i\in I0} \log P\{h_i=0\} + \Sigma_{i\in I1} [\log f(y_i|h_i=1) + \log P\{h_i=1\}]$$

- Two step approach (Heckman, 1979)
  1. Estimate the coefficients β_2 of the selection equation by standard probit maximum likelihood
  2. Compute $\lambda(x_{2i}'b_2) = \phi(x_{2i}'b_2) / \Phi(x_{2i}' b_2) = \lambda_i$
  3. Estimate the coefficients β_1 and σ_{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 Tobacco

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	
const	0,0444178	0,0492440	0,9020	0,3671	
AGE	0,00874370	0,0110272	0,7929	0,4278	
NADULTS	-0,0130898	0,0165677	-0,7901	0,4295	
NKIDS	-0,00221765	0,000585669	-3,787	0,0002	***
NKIDS2	-0,00260186	0,00228812	-1,137	0,2555	
LNX	-0,00174557	0,00357283	-0,4886	0,6251	
AGELNX	-0,000485866	0,000807854	-0,6014	0,5476	
NADLNX	0,000817826	0,00119574	0,6839	0,4940	
WALLOON	0,00260557	0,000958504	2,718	0,0066	***
lambda	-0,00013773	0,00291516	-0,04725	0,9623	

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 Tabacco, cont'd

Heckit ML estimation, GRETL output

Model 7: ML Heckit, using observations 1-2724
Dependent variable: SHARE1
Selection variable: D1

Selection equation

	coefficient	std. error	t-ratio	p-value	
const	-16,2535	2,58561	-6,286	3,25e-010	***
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-011	***
AGELNX	-0,0510698	0,0486730	-1,049	0,2941	
NADLNX	-0,160399	0,0748929	-2,142	0,0322	**
BLUECOL	-0,0352022	0,0983073	-0,3581	0,7203	
WHITECOL	0,0801599	0,0852980	0,9398	0,3473	
WALLOON	0,201073	0,0628750	3,198	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

	constant	NKIDS	LNX	WALL
Tobit model	-0,1704	-0,0030***	0,0134***	0,0042***
	0,0441	0,0008	0,0033	0,0010
Truncated regression	0,0433	-0,0022***	-0,0017	0,0026***
	0,0458	0,0008	0,0034	0,0009
Tobit II model	0,0444	-0,0022***	-0,0017	0,0026***
	0,0492	0,0006	0,0036	0,0010
Tobit II selection	-16,2535	-0,0936**	1,2583***	0,2011***
	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_0: $\sigma_{12} = 0$ in the second step of Heckit, i.e., fitting the regression $w_i = x_{1i}'\beta_1 + \sigma_{12}\lambda_i + \eta_i$
- *t*-test on the coefficient for Heckman's lambda
- Test results are sensitive to exclusion restrictions on x_{1i}

Tobit Models in GRETL

Model > Nonlinear Models > Tobit

- Estimates the Tobit model; censored dependent variable

Model > Nonlinear Models > Heckit

- Estimates in addition the selection equation (Tobit II), optionally by ML- and by two-step estimation

Your Homework

1. Verbeek‘s data set CREDIT contains credit ratings of 921 US firms, as well as characteristics of the firm; the variable *rating* has categories “1“, …,“7“ (highest) . Generate the variable GF (good firm) with value 1 if *rating* > 4 and 0 otherwise, and the more detailed variable CR (credit rating) with CR = 1 if *rating* < 3, CR = 2 if *rating* = 3, CR = 3 if *rating* = 4, and CR = 4 otherwise.
   a. Estimate a binary logit model for the assignment of the GF ratings, and an ordered logit model for assignment CR.
   b. Compare the effects of the regressors in the models, based on coefficients and slopes.
   c. What is the percentage of firms correctly rated by GF that are incorrectly rated by CR?
2. People buy for y_i* of an investment fund, with $y_i^* = x_i'\beta + \varepsilon_i$ with ε_I ~ N(0,1); x_i consists of an intercept and the variables age and income. The dummy $d_i = 1$ if $y_i^* > 0$ and $d_i = 0$ otherwise.

Your Homework, cont'd

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.

3. Verbeek's data set TOBACCO contains also expenditures on alcohol in 2724 Belgian households, taken from the Belgian household budget survey of 1995/96, as well as other characteristics of the households; for the expenditures on alcohol, the dummy D1=1 if the budget share for alcohol SHARE1 differs from 0, and D1=0 otherwise.
   a. Model the budget share for alcohol, using (i) a Tobit model, (ii) a truncated regression, and (iii) a Tobit II model, using besides other characteristics of the household the dummy FLANDERS.
   b. Compare the effects of the regressors in the models, based on coefficients and slopes.
   c. Compare the results for Flanders with that for the Wallonie.