Econometrics 2 - Lecture 2

Models with Limited Dependent Variables

Contents

Example

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 will 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 $y_i = x_i'\beta + \varepsilon_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i$

- 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 $E\{y_i|x_i\} = 1*P\{y_i = 1|x_i\} + 0*P\{y_i = 0|x_i\} = P\{y_i = 1|x_i\}$
- 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
    - ε_i can take on 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

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

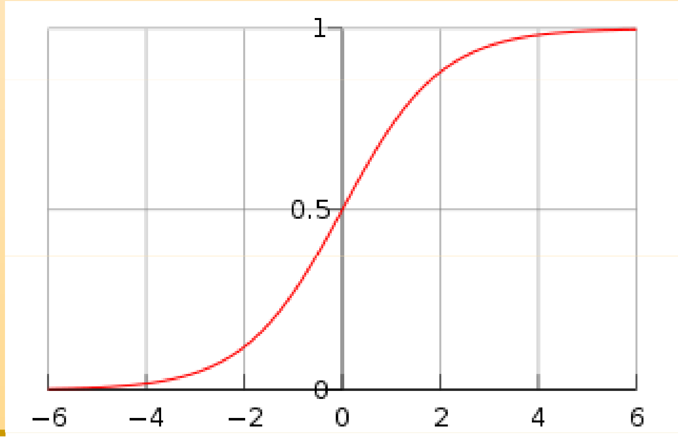
Car Ownership and Income, cont'd

E.g., P{y_i =1|x_i} = 1/(1+exp(-z_i)) with z = -0.5 + 1.1*x, the income 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

Standard logistic distribution function $L(z)$, with z on the horizontal and $L(z)$ on the vertical axis

x	z	P{y =1|x}
1	0.6	0.646
2	1.7	0.846
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 $p/(1-p)$ 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

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

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

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

- 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)

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

Linear Probability Model (LPM)

Assumes that

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

but sets restrictions

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

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

- 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

$E\{y_i|x_i\} = P\{y_i = 1|x_i\} = F(x_i'\beta)$: assume $F(.)$ to be the distribution function of $N(0, \sigma^2)$

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

- 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

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 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 of β in the logit model are roughly a factor $\pi/\sqrt{3} \approx 1.81$ 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 β_k

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

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

- The marginal effect of changing x_k
  - Probit model: $\phi(x_i'\beta)\,\beta_k$, with standard normal density function ϕ
  - Logit model: $\exp\{x_i'\beta\}/[1 + \exp\{x_i'\beta\}]^2\,\beta_k$
  - Linear probability model: β_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 β_k; the sign is that of β_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 $E\{y_i|x_i\}$ for 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 ratio when increasing x_k by 1 unit

Contents

Binary Choice Models: Estimation

Typically, binary choice models are estimated by maximum likelihood

Likelihood function, given N observations (y_i, x_i)

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

$$= \Pi_i F(x_i'\beta)^{y_i} (1 - F(x_i'\beta))^{1-y_i}$$

- Maximization of 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 $\Sigma_i e_i x_i = 0$ define the generalized residuals

$$e_i = \frac{y_i - F(x_i'\beta)}{F(x_i'\beta)(1-F(x_i'\beta))} f(x_i'\beta)$$

- The generalized residuals e_i can assume two values, depending on the value of y_i:
  - $e_i = f(x_i'b)/F(x_i'b)$ 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 [due to P{y_i =1|x_i} = p_i = $L(x_i,\beta)$]

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

- From $\Sigma_i \hat{p}_i x_i = \Sigma_i y_i x_i$ follows – given that the model contains an intercept –:
  - 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 > 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

Example: Effect of Teaching Method

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

Personalized System of Instruction: a 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 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

	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

Estimated logit model for the indicator GRADE

$$P\{GRADE = 1\} = p = L(z) = \exp\{z\}/(1+\exp\{z\})$$

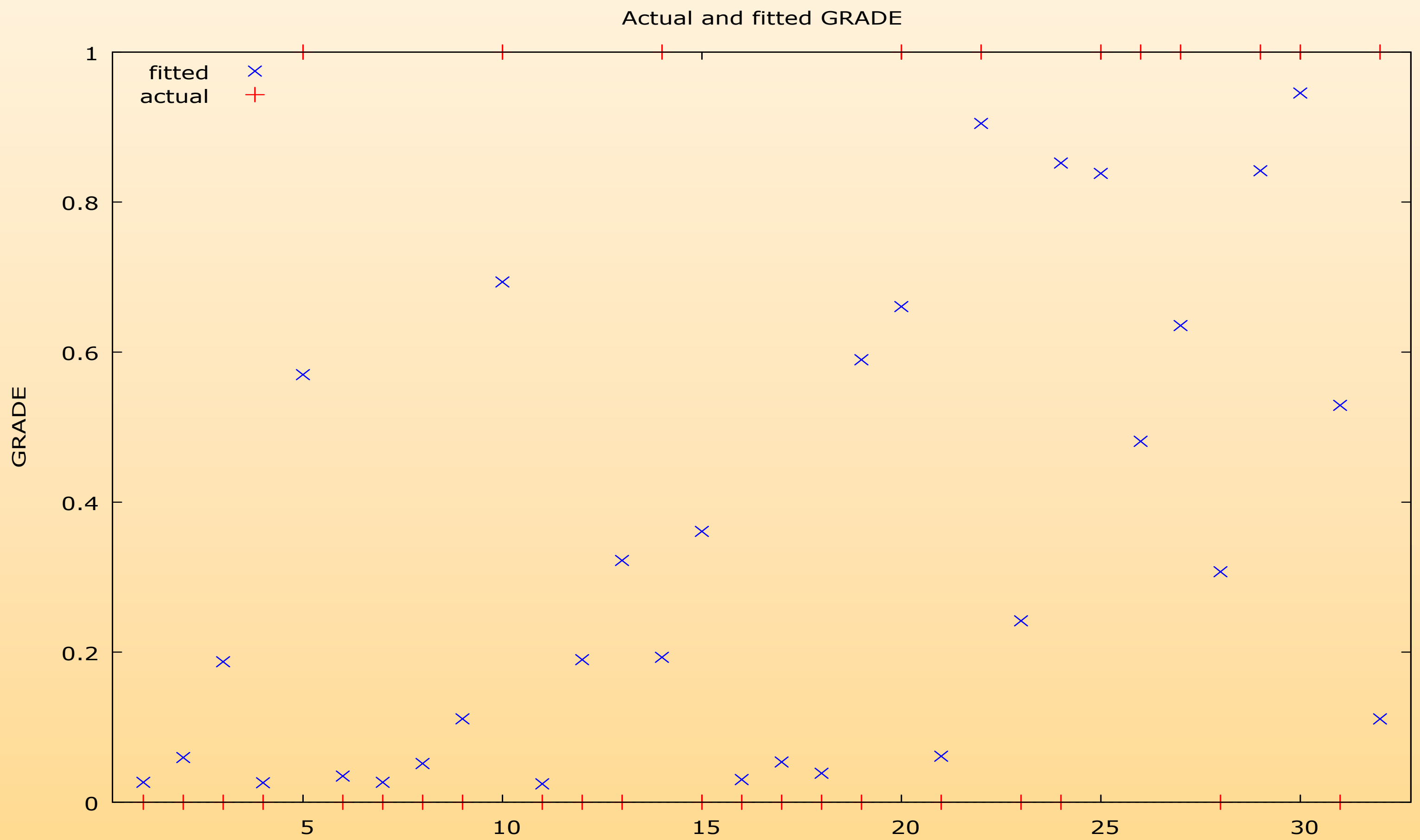
with

$$z = -13.02 + 2.826*GPA + 0.095*TUCE + 2.38*PSI$$
$$= \log\{p/(1-p)\} = \text{logit}\{p\}$$

- 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

Effect of Teaching Method, cont'd

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



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

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 or hit rates
  - R_p^2

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 $\ell_1 = \ell_0$, i.e., all slope coefficients are zero
- *McFadden* R^2 attains the upper limit 1 if $\ell_1 = 0$

Naïve Model: Calculation of ℓ_0

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

- $P\{y_i = 1\} = p$ 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_p^2

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 $F(x_i'b) < 0.5$, i.e., 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

    $\hat{y}_i = 1$ for all i (!), if $\hat{p} = N_1/N > 0.5$; $\hat{y}_i = 0$ for all i if $\hat{p} \leq 0.5$
  - $wr_0 = 1 - \hat{p}$ if $\hat{p} > 0.5$, $wr_0 = \hat{p}$ if $\hat{p} \leq 0.5$
  - Goodness-of-fit measure: $R_p^2 = 1 - wr_1/wr_0$; may be negative!

	$\hat{y} = 0$	$\hat{y} = 1$	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 pre-test, 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

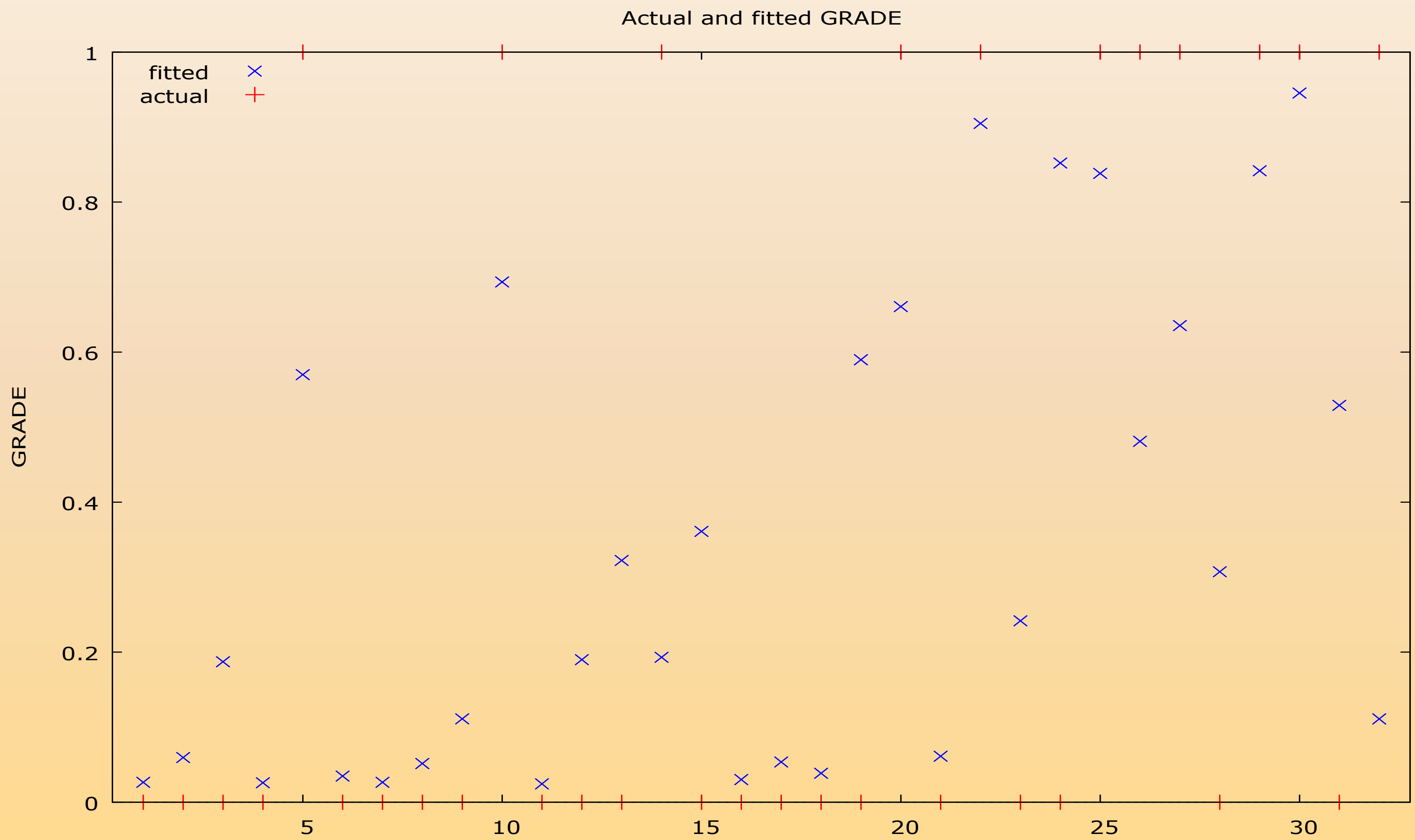
Mean dependent var	0.343750	S.D. dependent var	0.188902
McFadden R-squared	0.374038	Adjusted R-squared	0.179786
Log-likelihood	-12.88963	Akaike criterion	33.77927
Schwarz criterion	39.64221	Hannan-Quinn	35.72267

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	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: $\ell_1 = -12.89$, $wr_1 = 6/32$

Goodness-of-fit measures

- McFadden $R^2 = 1 – (-12.89)/(-20.59) = 0.374$
- *pseudo*-$R^2 = 1 - 1/[1 + 2(-12.89 + 20.59)/32) = 0.325$
- $R_p^2 = 1 – wr_1/wr_0 = 1 – 6/11 = 0.45$

Contents

Modelling Utility

Latent variable y_i^*: 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_i^* depends upon observed characteristics (e.g., 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 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

- ε_i's are independent of x_i's
- ε_i has a 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

Contents

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

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

- M+1 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 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

$$y_i = 1 \text{ if } y_i^* \le 0$$
$$y_i = 2 \text{ if } 0 < y_i^* \le \gamma$$
$$y_i = 3 \text{ if } y_i^* > \gamma$$

- 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:

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

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

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

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

Interpretation of parameters β

- Wrt $y_i^*(= x_i'\beta + \varepsilon_i)$: willingness to work increases with larger x_k for positive β_k
- Wrt probabilities $P\{y_i = j | x_i\}$, e.g., for positive β_k
  - $P\{y_i = 3 | x_i\} = P\{y_i^* > \gamma | x_i\}$ 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:

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			Ordered logit	
	Estimate	Standard error		Estimate	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

Latent variable model

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

with explanatory variables x_i

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

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

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

Contents

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, $j = 1,\ldots, M$,
- chooses the alternative with the highest utility level $\max\{U_{i1}, ..., U_{iM}\}$

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

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

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

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

  for $j = 1, \ldots, M$
- and $\Sigma_j P\{y_i = j\} = 1$
- 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 $j = 1, \ldots, M$

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

- Alternative-specific characteristics x_{ij}
- E.g., mode of transportation (by car, train, bus) is affected by the travel costs, travel time, etc. of the individual *i*

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

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

- 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 for the mixed logit model: it combines
  - alternative-specific characteristics and
  - person-specific characteristics

Number of parameters

- conditional logit model: vector β with K components
- multinomial logit model: vectors β_2, ..., β_M, each with K components

Independence of Errors

Independence of the error terms ε_{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 > 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...`

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

Contents

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 a patent, as a function of
  - Firm size
  - R&D expenditures
  - Industrial sector
  - Country, etc.

See Verbeek's data set PATENT

Example: Patents and R&D Expenditures

Verbeek's data set PATENTS: number of patents (p91), expenditures for R&D (logrd91), 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_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

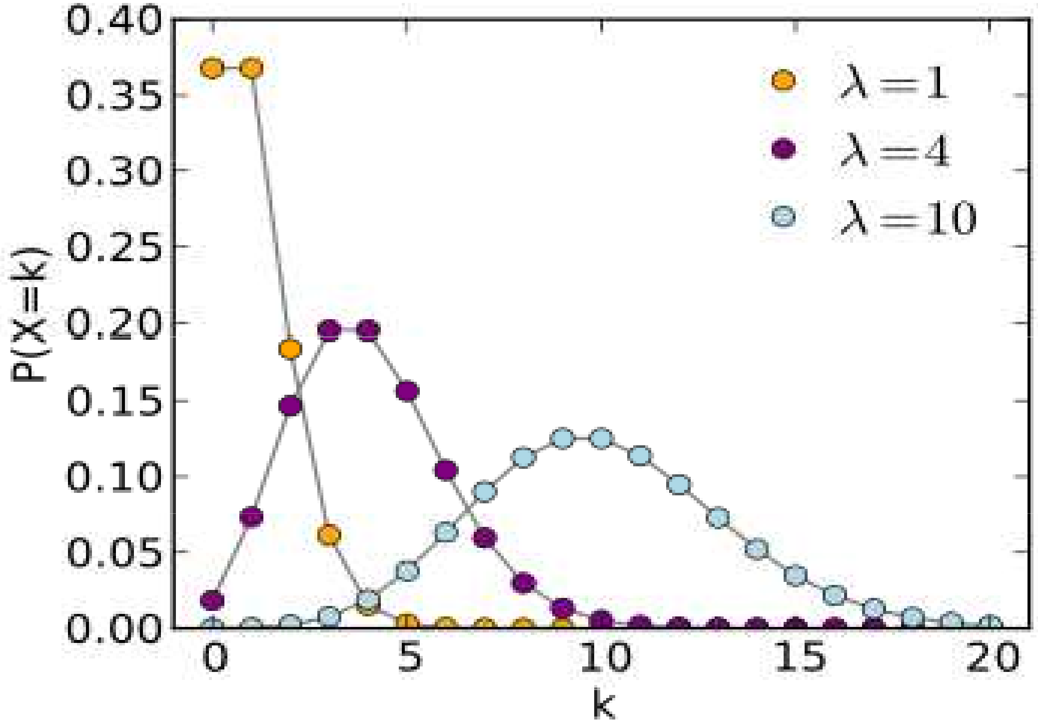
$$P\{y_i = y \mid x_i\} = \frac{\lambda_i^y}{y!}\exp\{\lambda_i\},\ y = 0, 1, \ldots$$

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

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



Poisson Regression Model: Estimation

Unknown parameters: coefficients β

Estimates of β allow assessing how $\exp\{x_i'\beta\} = E\{y_i \mid x_i\}$ is affected by x_i

Fitting the model to data: ML estimators for β are

- Consistent
- Asymptotically efficient
- Asymptotically normally distributed

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?

Model:

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

- y_i: number of patents in company i in year 1991
- x_i: 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	−4950.789	Akaike criterion	9919.578
Schwarz criterion	9948.365	Hannan-Quinn	9931.249

Overdispersion test: Chi-square(1) = 18.6564 [0.0000]

Poisson Regression Model: Overdispersion

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

Count Data Models in GRETL

`Model > Nonlinear Models > 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

Tobit Models

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

Examples:

- 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
- Expenditures on durable goods as a function of income, age, etc.: a part of units does not spend any money on durable goods

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:

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

- y_i^*: optimal expenditures on tobacco in household *i* (latent)
- x_i: characteristics of the *i*-th household
- ε_i: unobserved heterogeneity (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

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

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

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

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 β_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 $E\{y_i \mid y_i > 0\}$ is different from β_k
- 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 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 (due to $y_i^* = x_i'\beta + \varepsilon_i$)

Contents

The Standard Tobit Model: Estimation

OLS produces inconsistent estimators for β; alternatives:

1. ML estimation based on the log-likelihood

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

   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:

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

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

Example: Model for Budget Share for Tobacco 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: $w_i^* = x_i'\beta + \varepsilon_I$

x_i: log of total expenditures (LNX) and various characteristics like

- number of children ≤ 2 years old (NKIDS2)
- number of adults in household (NADULTS)
- Age (AGE)

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

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

- 2724 households

Model for Budget Share

Budget share w_i^* for expenditures on alcoholic beverages

$$w_i^* = x_i'\beta + \varepsilon_I$$

regressors x_i:

- 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)

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

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. 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 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)

	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

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
- for omitted variables

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 to the estimation of σ^2

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

Contents

An Example: Modeling Wages

Wage observations: available only for the working population

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

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

  with exogenous characteristics (age, education, …)
- Selection equation or labor force participation

  $h_i^* = x_{2i}'\beta_2 + \varepsilon_{2i}$
- Observation rule: w_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}: usually normality with

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

Model for Wages: Selection Equation

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

- 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}
  - x_{2i} should contain variables not included in x_{1i}
  - Sign and value of coefficients of the same variables in x_{1i} and x_{2i} are not the same
- Special cases
  - If $\sigma_{12} = 0$, sample selection is exogenous
  - Tobit II model coincides with Tobit I model if $x_{1i}'\beta_1 = x_{2i}'\beta_2$ and $\varepsilon_{1i} = \varepsilon_{2i}$

Model for Wages: 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 error, consistent OLS estimates of the wage equation

Tobit II Model: Log-likelihood Function

Log-likelihood

$$\ell_3(\beta_1,\beta_2,\sigma_1^2,\sigma_{12}) = \Sigma_{i\in I0}\log P\{h_i=0\} + \Sigma_{i\in I1} [\log f(y_i|h_i=1)+\log P\{h_i=1\}]$$

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

with

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

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

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

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

  $\ell_3(\beta_1,\beta_2,\sigma_1^2,\sigma_{12}) = \Sigma_{i\in I0} \log P\{h_i=0\}+\Sigma_{i\in I1} [\log f(y_i|h_i=1)+\log P\{h_i=1\}]$
- Two step approach (Heckman, 1979)
  1. Estimate the coefficients β_2 of the selection equation by standard probit maximum likelihood: b_2
  2. Compute estimates of Heckman's lambdas: $\lambda_i = \lambda(x_{2i}'b_2) = \phi(x_{2i}'b_2) / \Phi(x_{2i}' b_2)$ for $i = 1, \ldots, N$
  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 Alcohol

Heckit ML estimation, GRETL output

D1: dummy,
1 if SHARE1 > 0

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 Alcohol, 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

		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 model	coeff.	0,0444	**-0,0022**	-0,0017	**0,0026**
	s.e.	0,0492	0,0006	0,0036	0,0010
Tobit II selection	coeff.	-16,2535	**-0,0936**	**1,2583**	**0,2011**
	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_0: $\sigma_{12} = 0$ in the second step of Heckit, i.e., fitting the regression $w_i = x_{1i}'\beta_1 + \sigma_{12}\lambda_i + \eta_i$
- GRETL: *t*-test on the coefficient for Heckman's lambda
- GRETL: Heckit-output shows rho, estimate for ρ_{12} from $\sigma_{12} = \rho_{12}\sigma_1$
- 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. 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 consists of a "1" for the intercept and the variable *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 ML estimator of the probability for $d_i = 1$ as function of x_i of the logit model.
2. 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.

Your Homework, cont'd

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 BLUECOL 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.