University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

9. Discrete Choice Models

Prof. Dr. Miroslav Verbič

miroslav.verbic@ef.uni-lj.si

www.miroslav-verbic.si

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Ljubljana, October 2022

Motivation

So far, we have focused on modelling continuous dependent variables, but in real life the number of alternatives is often small. This requires a particular modelling approach.

Discrete choice models → the variable to be explained, *y*, is taking a small finite number of outcomes; i.e. we have a discrete dependent variable.

↳ only a couple possible values

Discrete choice models

The **discrete dependent variable** can be:

- Binary or binomial: dichotomous choice [0, 1].

  Examples: work/not work, buy/not buy etc.
  Approach: **probit model & logit model**.
- Multinomial: multiple choice, e.g. [1, 2, 3, 4], which can be ordered or non-ordered.

  ↳ more than 2

  Examples:
  - non-ordered: mode of transport etc;
  - ordered: survey scale etc. → natural order (1 not pleased 5 very pleased)

  Approach:
  - non-ordered: **multinomial probit & logit model**;
  - ordered: **ordered probit & logit model**.

Discrete choice models

- Non-negative integer: count data [0, 1, 2, …].

  Examples: number of patents, number of loss events etc.
  Approach: **Poisson model & negative binomial model**.

We will first deal with the **binary** dependent variable.

0, 1 y

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

9.1 Binomial Discrete Choice Models

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Binomial discrete choice models

We are interested in the conditional or response **probability**:

$P(y = 1|X) = P(y = 1|x_1, \dots, x_k)$ for various values of x_j.

Why not just use a linear regression model, called the **linear probability model** (LPM), for the binary response variable *y*:

$$P(y_i = 1|x_i) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} \ ,$$

estimate it by the ordinary least squares estimator (OLS) and obtain the partial effects (regression coefficient estimates):

$$\beta_k = \frac{\partial P(y_i=1|x_i)}{\partial x_k} \ ?$$

Binomial discrete choice models

The distribution of the dependent variable is not normal ⇒ that is why we need to use this

Bernoulli–type random variables *y* and (thus) *u*:

- $P(y = 1|x) = p(x)$
- $P(y = 0|x) = 1 - p(x)$
  - $\mathrm{E}(y|x) = p(x)$
  - $\mathrm{Var}(y|x) = p(\mathrm{x})(1-p(\mathrm{x}))$
- $y \sim B(p(x), p(\mathrm{x})(1-p(\mathrm{x})))$

  ↳ not a normal distribution

Binomial discrete choice models

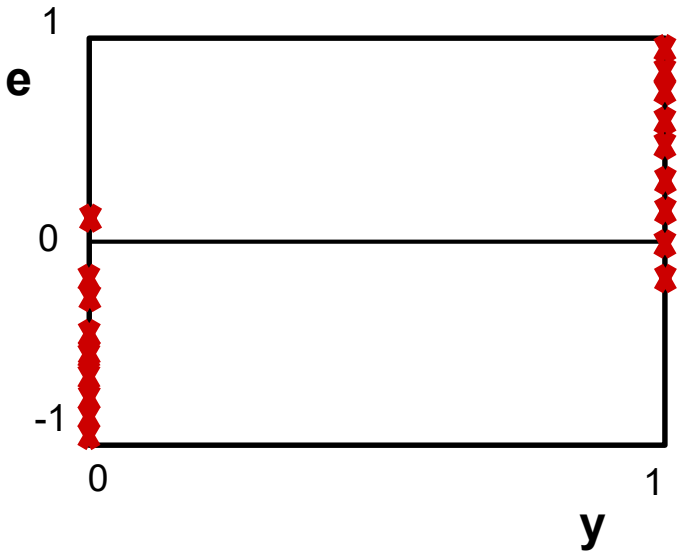
Therefore, we have several **reasons**:

1) Non-normality of the random variable *u* (and *y*);
2) Heteroscedasticity of the variance of the random variable (variance depends on x_j);
3) Predicted probabilities, i.e. fitted values $\hat{y}$, could lie outside the unit interval $0 \leq E(y_i|x_{ji}) \leq 1$, which could further imply negative variance (for $\hat{y} < 0$; see previous slide);
4) Questionable power of R^2, since all residuals are concentrated at only two values of *y* (see next slide);
5) Questionable linear relationship between *y* and *x* for such a model; are constant (marginal) effects of *x* on *y* realistic?

they can be negative

Binomial discrete choice models

All the residuals are concentrated in 1 and 0.



We usually use a **probit** or a **logit model** instead, which is estimated by the **maximum likelihood estimator** (MLE).

└ the second most widely used estimator, we use it instead of the OLS estimator

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

9.1.1 Maximum Likelihood Estimation

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Basic concepts

- ✓Maximum likelihood estimation (MLE) is a statistical method (an estimator) to find the most likely density function that would have generated the data.
- ✓Thus, MLE requires you to make a distributional assumption first.
- ✓We will provide the intuition behind the MLE using some examples.

Basic concepts

Id	x
1	1
2	4
3	5
4	6
5	9

- ✓ Let us explain the basic idea of MLE using the data on the left.
- ✓ Let us make an assumption that the variable x follows normal distribution.
- ✓ Remember that the density function of normal distribution with mean μ and variance σ^2 is given by:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \quad \text{for} \quad -\infty < x < \infty$$

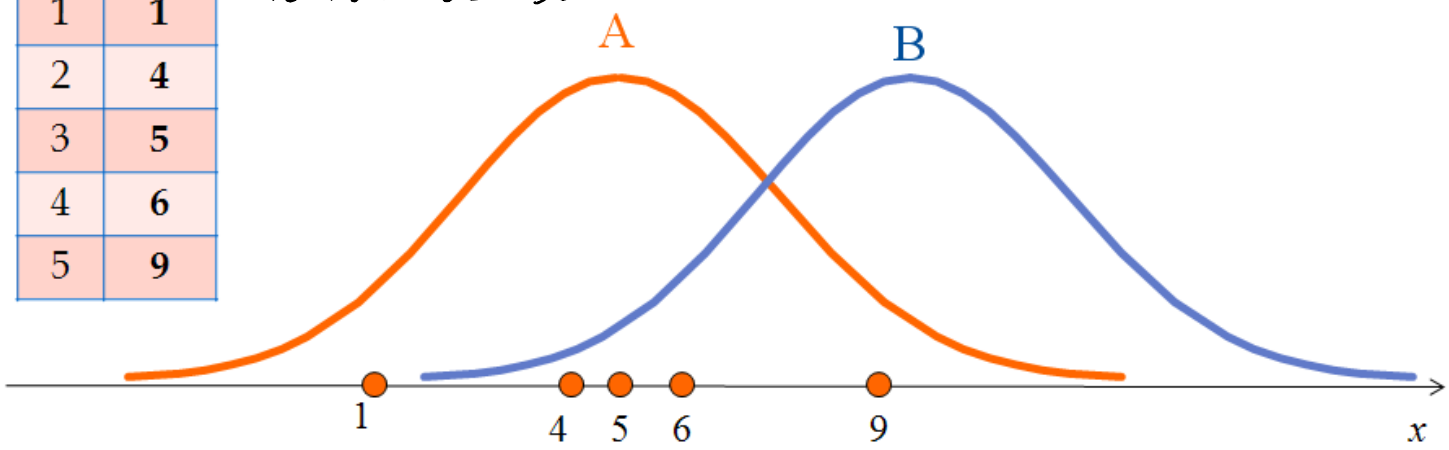
Basic concepts

- ✓The data are plotted on the horizontal line.
- ✓Now, ask yourself the following question:

"Which distribution, A or B, is more likely to have generated the data?"

Id	x
1	1
2	4
3	5
4	6
5	9

A because the density is clustered around it.



Basic concepts

- ✓Answer to the question is A, because the data are clustered around the center of the distribution A, but not around the center of the distribution B.
- ✓This example illustrates that, by looking at the data, it is possible to find the distribution that is most likely to have generated the data.
- ✓Now, how exactly do we find the distribution in practice?

Estimation procedure in general

- ✓MLE starts with computing the likelihood contribution of each observation.
- ✓The likelihood contribution is the height of the density function. We use L_i to denote the likelihood contribution of i^{th} observation.

Estimation procedure in general

Graphical illustration of the likelihood contribution:

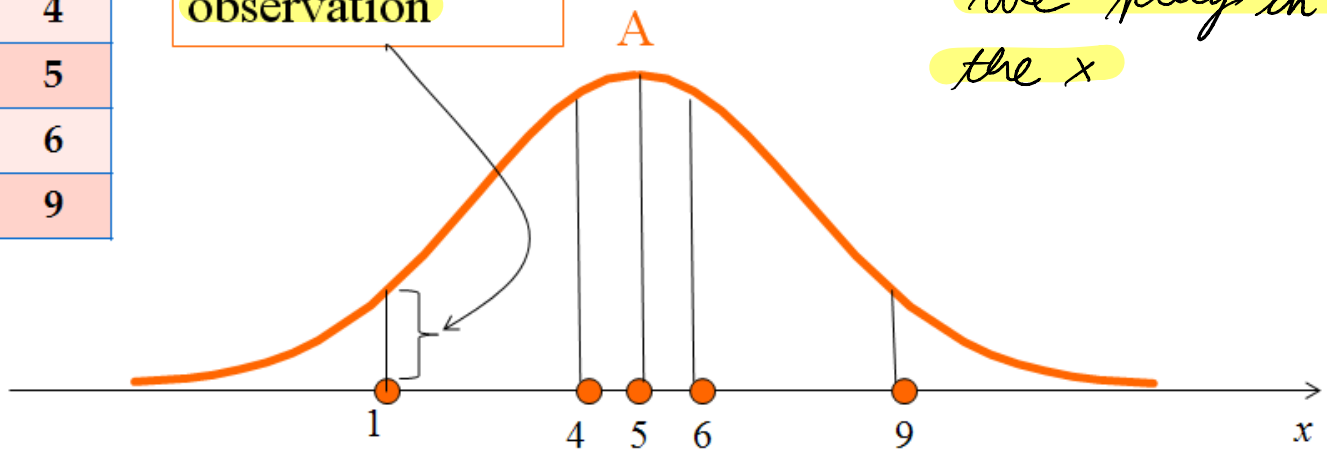
Id	x
1	1
2	4
3	5
4	6
5	9

The likelihood contribution of the first observation

$$= L_1 = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(1-\mu)^2/2\sigma^2}$$

Data value

we plug in the x



Estimation procedure in general

✓Then, we multiply the likelihood contributions of all the observations. This is called the likelihood function, L:

$$L = \prod_{i=1}^{n} L_i$$

This notation means you multiply from $i = 1$ through n.

= product

✓In our example, $n = 5$.

Estimation procedure in general

- In our example, the likelihood function looks like:

Id	x
1	1
2	4
3	5
4	6
5	9

$$L(\mu,\sigma)=\prod_{i=1}^{5} L_i = L_1 \times L_2 \times L_3 \times L_4 \times L_5$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(1-\mu)^2/2\sigma^2} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(4-\mu)^2/2\sigma^2}$$

$$\times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(5-\mu)^2/2\sigma^2} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(6-\mu)^2/2\sigma^2}$$

$$\times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(9-\mu)^2/2\sigma^2}$$

- We write $L(\mu, \sigma)$ to emphasize that the likelihood function depends on these parameters.

Estimation procedure in general

- ✓Then we find the values of μ and σ that maximize the likelihood function.
- ✓The values of μ and σ which are obtained this way are called the maximum likelihood estimates (MLEs) of μ and σ.
- ✓Most of the MLEs cannot be solved 'by hand'. Thus, we need to apply an iterative procedure to solve it on computer.
- ✓Fortunately, the majority of models that require MLE can be estimated automatically in Stata and R.

Estimation of a regression function

- ✓We are usually interested in estimating a linear regression function.
- ✓We will use a simple bivariate regression model for illustration:

$$y = \beta_0 + \beta_1 x + u.$$

- ✓Estimation of such a model can be done using the MLE.

Estimation of a regression function

Id	y	x
1	2	1
2	6	4
3	7	5
4	9	6
5	15	9

- ✓ Suppose that we have these data, and we are interested in estimating the above model.
- ✓ Let us make an assumption that u follows the normal distribution with mean 0 and variance σ^2.

Estimation of a regression function

- We can rewrite the model as:

$$u = y - (\beta_0 + \beta_1 x).$$

- This means that $y - (\beta_0 + \beta_1 x)$ follows the normal distribution with with mean 0 and variance σ^2.
- The likelihood contribution of each observation is the height of the density function at the data point $y - (\beta_0 + \beta_1 x)$.

this is now a data point

Estimation of a regression function

For example, the likelihood contribution of the 2nd observation is given by:

Id	y	x
1	2	1
2	6	4
3	7	5
4	9	6
5	15	9

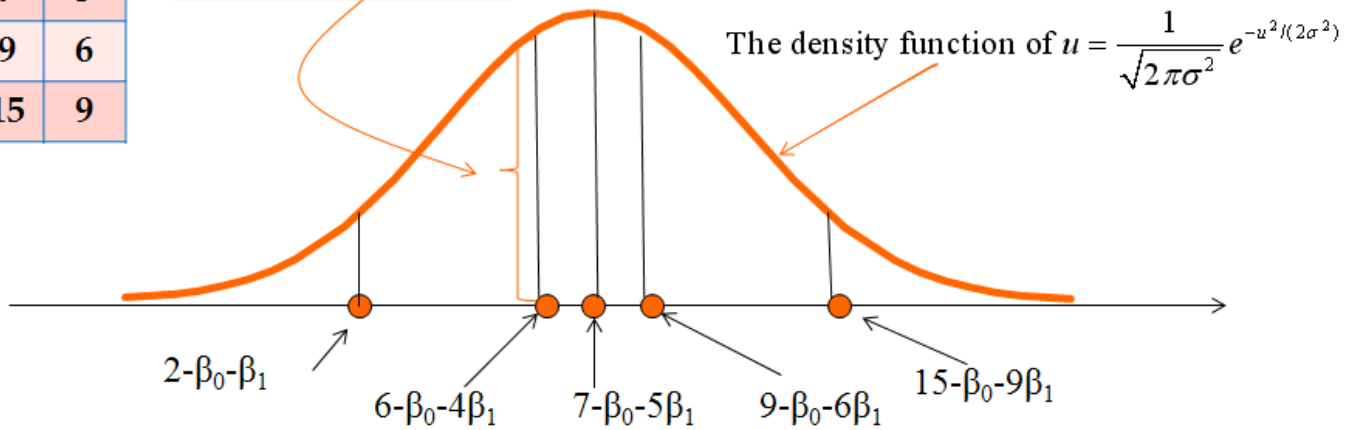
The likelihood contribution of the 2nd observation

Data point

$$=L_2 = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(6-\beta_0-4\beta_1)^2/2\sigma^2}$$

$u_2-\mu=(y_2-\beta_0-\beta_1 x_2)-0$

The density function of $u = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-u^2/(2\sigma^2)}$



Estimation of a regression function

Then the likelihood function is given by:

Id	y	x
1	2	1
2	6	4
3	7	5
4	9	6
5	15	9

$$L(\beta_0, \beta_1, \sigma) = \prod_{i=1}^{n} L_i = L_1 \times L_2 \times L_3 \times L_4 \times L_5$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(2-\beta_0-\beta_1)^2/2\sigma^2} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(6-\beta_0-4\beta_1)^2/2\sigma^2}$$

$$\times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(7-\beta_0-5\beta_1)^2/2\sigma^2} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(9-\beta_0-6\beta_1)^2/2\sigma^2}$$

$$\times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(15-\beta_0-9\beta_1)^2/2\sigma^2}$$

Estimation of a regression function

- ✓ The likelihood function is thus a function of β_0, β_1, and σ.
- ✓ We choose the values of β_0, β_1, and σ that maximize the likelihood function. These are the maximum likelihood estimates of of β_0, β_1, and σ.
- ✓ Again, maximization can easily be done automatically in Stata and R.
  - ↳ we still need to understand what is going on.

Summary of the MLE procedure

1. Compute the likelihood contribution of each observation, L_i, for i = 1,…,n.
2. Multiply all the likelihood contributions to form the likelihood function, L:

$$L = \prod_{i=1}^{n} L_i$$

3. Maximize L by choosing the values of the parameters. The values of parameters that maximize L are the maximum likelihood estimates of the parameters.

Properties of the ML estimates

most important property

1. Consistency
2. Asymptotic normality

if the sample is large enough we have normality

3. Asymptotic efficiency

if the sample is large enough we have efficiency

4. Invariance

Invariance means that for any one-to-one transformation of the model parameters, the ML estimates (the maximization solution) remain unchanged.

ML estimator is used on large sample

The log-likelihood function

- It is usually easier to maximize the natural log of the likelihood function than the likelihood function itself:

$$\ln(L) = \ln\left[\prod_{i=1}^{n} L_i\right] = \ln(L_1 \cdot L_2 \cdot \ldots \cdot L_n) =$$

$$= \ln(L_1) + \ln(L_2) + \ldots + \ln(L_n) = \sum_{i=1}^{n} \ln(L_i)$$

- Due to invariance, maximizing the so called log-likelihood function is identical to maximizing the likelihood function.

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

9.1.2 Latent Variable Approach

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Example from real life

The university would like to evaluate your **knowledge** at the end of each course.

Unfortunately, the knowledge is not (directly) observed, only your **exam results** can be evaluated, which is not always the same thing as obtained knowledge.

At the doctoral/PhD level, often the only two grades awarded are pass and fail (no grades from 1 to 10).

Example from economics

Labour force participation, LFP:

= 1, if an individual participates in the labour market (works) or

= 0, if he/she does not participate in the labour market.

Rational individual maximizes his **direct utility function**, subject to his **budget constraint**:

$$\max\ u(c,j),\quad \text{s.t.}\ \ y_N + w(H-j) = c$$

where *c* stands for consumption of goods, *j* for consumption of leisure time, y_N for non-labour income, *w* for wage rate, and *H* for total available time.

Example from economics

We derive:

- the indirect utility of inactivity: $v(H, y_N)$ and
- the indirect utility of activity: $v(w, H, y_N)$.

Note that apart from w, H *and* y_N, everything else is endogenous.

The following holds for a **rational individual**:

$LFP = 1$ if and only if $v(w, H, y_N) \geq v(H, y_N)$, and 0 otherwise.

indirect utility of activity must be higher than indirect utility of inactivity

Latent variable approach

We often do not observe the underlying **choice variables** (e.g. indirect utility), but we do observe the **choice itself** (e.g. LFP).

We assume (by rationality) that the option with more favourable choice variable value was chosen.

Latent variable approach

In econometrics, we model this by the so called **latent-variable approach**:

$$y_i^* = X\beta + u_i \ ,$$

where *y** is the latent (unobserved) variable (e.g. *v*) and the following observed outcomes (on *y*, e.g. on *LFP*) with assumed relationships (about *y**, e.g. about *v*) hold:

$$y_i = 1, \ \ if \ y_i^* \geq 0;$$
$$y_i = 0, \ \ if \ y_i^* < 0.$$

we're going to work if it pays off
we're not going to work if it doesn't pay off.

we can change the sign so that

$$y = 1, \text{ if } y_i^* \leq 0$$
$$y = 0, \text{ if } y_i^* > 0$$

Latent variable approach

We model the probability of a choice:

We are modeling the probability of the stochastic variable

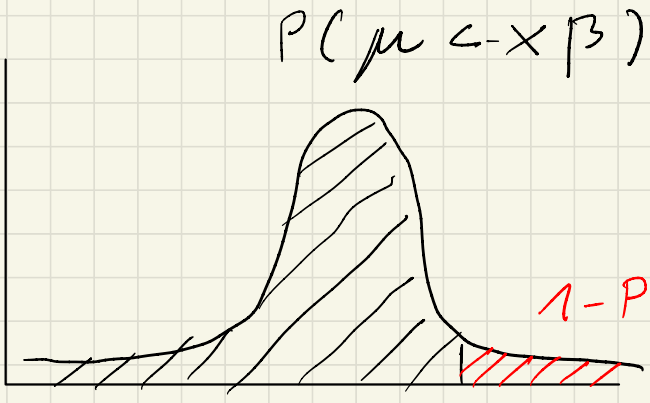
$$P(y=0) = P(y^* < 0) = P(X\beta + u < 0) = P(u < -X\beta) =$$
$$= \Psi(-X\beta) = 1 - \Psi(X\beta)$$

because of symmetry

and

$$P(y=1) = P(y^* \geq 0) = P(X\beta + u \geq 0) = P(u \geq -X\beta) =$$
$$= 1 - \Psi(-X\beta) = \Psi(X\beta),$$

where $\Psi(\cdot)$ is the cumulative distribution function (CDF) and it holds that $\Psi(X\beta) + \Psi(-X\beta) = 1$ (symmetry).


P(μ <- x β)
1 - P

Latent variable approach

For binary choice, the probability of an observation with outcome either $y_i = 0$ or $y_i = 1$ is:

$$P(y_i|X) = (\Psi(X\beta))^{y_i} \cdot (1 - \Psi(X\beta))^{1-y_i} = L_i(\beta),$$

which is called the **likelihood contribution** of observation *i*, L_i.

Usually, we utilize the **log-likelihood contribution** of observation *i*, denoted by ln L_i:

$$lnL_i(\beta) = y_i ln\Psi(X\beta) + (1 - y_i) ln(1 - \Psi(X\beta)).$$

Latent variable approach

For **maximum likelihood estimation** of **β**, we need to assume a form for the cumulative distribution function, Ψ(Xβ).

For binary choice, we have two possibilities:

1. Standard normal distribution:

$$\Psi(X\beta) = \Phi(X\beta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{X\beta} e^{-\frac{1}{2}t^2} dt, \qquad t \sim N(0,1)$$

leads to the **probit model**;

Latent variable approach

2. Logistic distribution:

$$\Psi(X\beta) = \Lambda(X\beta) = \frac{e^{X\beta}}{1+e^{X\beta}} = \frac{1}{\frac{1}{e^{X\beta}}+1} = \frac{1}{1+e^{-X\beta}}$$

leads to the **logit model**.

Conveniently (as we model probabilities), for both cumulative distribution functions it holds that:

$$0 \leq \Phi(X\beta) \leq 1;$$
$$0 \leq \Lambda(X\beta) \leq 1.$$

Back to the example from real life…

unobserved – Latent variable: y^* – obtained knowledge

Observed variable – exam result, y:

$$y_i = \begin{cases} 1, if\ y_i^* \geq y_{min}\ (pass) \\ 0, if\ y_i^* < y_{min}\ (fail) \end{cases}$$

Implicit assumption: exams were fair in terms of:

a) no cheating and

b) fair grading.

Back to the example from real life…

What are the determinants of exams results?

$$y_i = f(age_i, H_i, D_i, E_i), \quad \forall \text{ student } i$$

where:

- y = 1 if passes, 0 if fails;
- *age* – age of a student;
- H – hours of studying the course;
- D – finished previous degree abroad (1 if yes, 0 if no);
- E – years of work experience.

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

9.1.3 The Probit Model

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Probit model and estimation

Let us assume that we have the following model:

$$y_i^* = \beta_0 + \beta_1 x_i + u_i$$

$$\begin{cases} \text{If } y_i = 0, \text{ then } y_i^* < 0 \\ \text{If } y_i = 1, \text{ then } y_i^* \geq 0 \end{cases}$$

Id	y	x
1	0	1
2	0	4
3	1	5
4	1	6
5	1	9

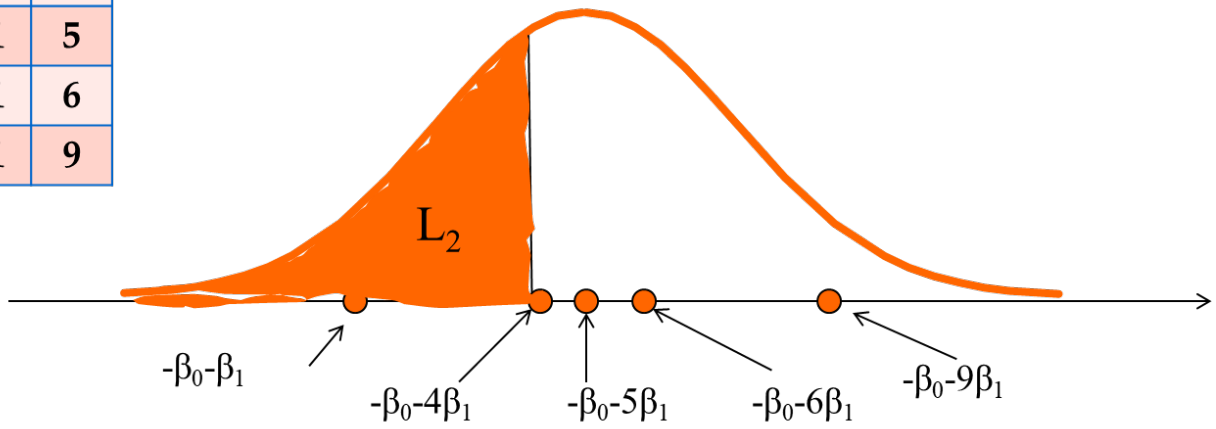
- ✓Suppose that we have the data on the left.
- ✓We also assume that $u_i \sim N(0,1)$.

Probit model and estimation

- ✓Take 2nd observation as an example. Since y=0 for this observation, we know y^*<0.
- ✓Thus, the likelihood contribution is:

Id	y	x
1	0	1
2	0	4
3	1	5
4	1	6
5	1	9

$$L_2 = P(y_2^* < 0) = P(\beta_0 + 4\beta_1 + u_2 < 0)$$
$$= P(u_2 < -\beta_0 - 4\beta_1) = \underbrace{\Phi(-\beta_0 - 4\beta_1)}_{\text{Cumulative distribution function of standard normal distribution}}$$

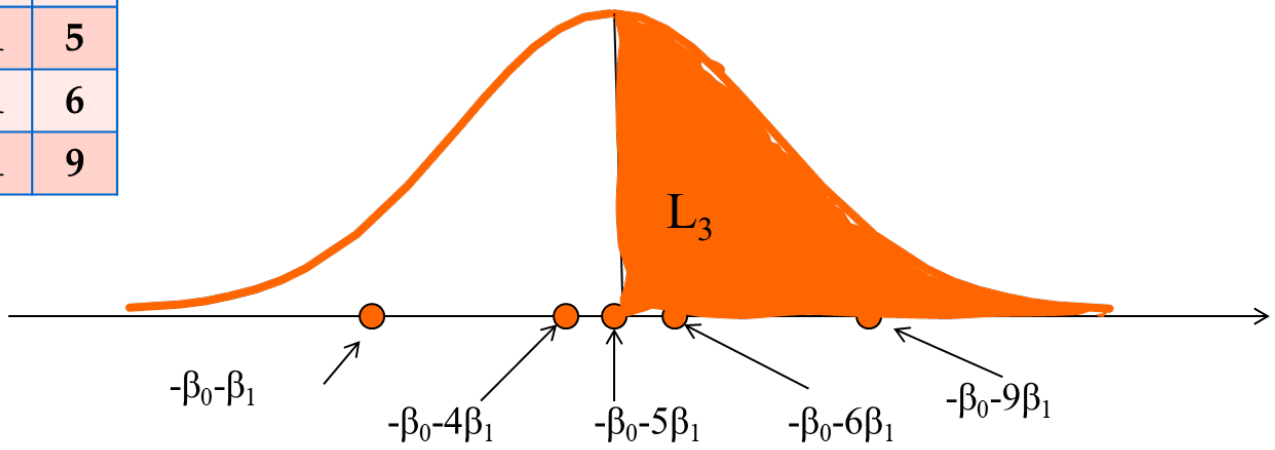


Probit model and estimation

- ✓Now, take 3rd observation as an example. Since y=1 for this observation, we know $y^*\geq 0$.
- ✓Thus, the likelihood contribution is:

Id	y	x
1	0	1
2	0	4
3	1	5
4	1	6
5	1	9

$$L_3 = P(y_3^* \geq 0) = P(\beta_0 + 5\beta_1 + u_3 \geq 0)$$
$$= P(u_3 \geq -\beta_0 - 5\beta_1) = 1 - \Phi(-\beta_0 - 5\beta_1)$$



Probit model and estimation

✓Then the likelihood function is given by:

Id	y	x
1	0	1
2	0	4
3	1	5
4	1	6
5	1	9

$$L(\beta_0,\beta_1)=\prod_{i=1}^{5}L_i=\Phi(-\beta_0-\beta)\Phi(-\beta_0-4\beta)\left[1-\Phi(-\beta_0-5\beta)\right]\times$$
$$\times\left[1-\Phi(-\beta_0-6\beta)\right]\left[1-\Phi(-\beta_0-9\beta)\right]$$

Probit model and estimation

- ✓Usually, we maximize the ln(L) instead of the L. Due to invariance, the result is identical.
- ✓The values of the parameters that maximize ln(L) are the ML estimators of the (binomial) *probit* model (sometimes it is also being called the *normit* model).
- ✓The MLE is done automatically in Stata and R. ↳ we are not going to do this manually

Generalization of the MLE procedure

The likelihood and the log-likelihood function:

$$L = \prod_{i=1}^{n} L_i(\beta) = \prod_{i=1}^{n} (\Phi(X\beta))^{y_i} \cdot (1 - \Phi(X\beta))^{1-y_i}$$

$$lnL = \sum_{i=1}^{n} l_i(\beta) = \sum_{i=1}^{n} y_i ln\Phi(X\beta) + \sum_{i=1}^{n} (1 - y_i) ln(1 - \Phi(X\beta))$$

The latter is being differentiated:

$\frac{\partial lnL}{\partial x_k} \Rightarrow \hat{\beta}_{k,ML}$ by using numerical methods (iterative procedures).

Marginal or partial effects

we can't interpret x the way we used to

The probit model is non-linear, therefore the estimated coefficients $\hat{\boldsymbol{\beta}}$ do not reflect the "strength" of the effects of a change in *x* on the probability of occurrence of *y*.

Instead, we calculate the **marginal effects**:

Probability density function of the std. normal distr.

- For a continuous x_k: $mfx_k = \frac{\partial p(x)}{\partial x_k} = \phi(X\beta)\beta_k$.

1. Since Φ(·) is strictly increasing, sign of β_k is the same as the sign of $\frac{\partial p(x)}{\partial x_k}$;

2. Relative effects do not depend on *x*: $\frac{\frac{\partial p(x)}{\partial x_k}}{\frac{\partial p(x)}{\partial x_j}} = \frac{\phi(X\beta)\beta_k}{\phi(X\beta)\beta_j} = \frac{\beta_k}{\beta_j}$.

Φ - cdf

ϕ - probability density function

Marginal or partial effects

for example: a dummy

- For a discrete x_k: a change of x_k from c to c+1 results in:

$$\Delta p(x) = \Phi\big(\beta_0 + \beta_1 x_1 + \cdots + \beta_k (c_k + 1)\big) - \Phi(\beta_0 + \beta_1 x_1 + \cdots + \beta_k c_k).$$

For a dummy explanatory variable x_k, c = 0:

$$\Delta p(x) = \Phi(\beta_0 + \beta_1 x_1 + \cdots + \beta_k \cdot 1) - \Phi(\beta_0 + \beta_1 x_1 + \cdots + \beta_k \cdot 0).$$

As you will see later, we usually evaluate marginal effects at mean values of all explanatory variables, $\phi(\bar{X}\beta)$, but in general, we can choose any values of our explanatory variables.

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

9.1.4 The Logit Model

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Logit model and estimation

- Again, consider the following model:

$$y_i^* = \beta_0 + \beta_1 x_i + u_i$$ — unobserved

$$\begin{cases} \text{If } y_i = 0, \text{ then } y_i^* < 0 \\ \text{If } y_i = 1, \text{ then } y_i^* \geq 0 \end{cases}$$ — observed

- In the (binomial) logit model, we assume that u_i follows the logistic distribution with mean 0 and variance 1, which has the following density function:

$$f(x) = \frac{e^{-x}}{1 - e^{-x}}$$

Logit model and estimation

Id	y	x
1	0	1
2	0	4
3	1	5
4	1	6
5	1	9

- ✓Now, suppose that you have the data on the left.
- ✓Take the 2nd observation as an example. Since $y=0$, it must have been the case that $y^*<0$.
- ✓Thus, the likelihood contribution is:

we are not going to do this manually

$$L_2 = P(y_2^* < 0) = P(\beta_0 + 4\beta_1 + u_2 < 0)$$

$$= P(u_2 < -\beta_0 - 4\beta_1) = \underbrace{\Lambda(-\beta_0 - 4\beta_1)}_{\text{Cumulative distribution function of logistic distribution}}$$

$$= \frac{1}{1+e^{-(-\beta_0-4\beta_1)}} = \frac{1}{1+e^{\beta_0+4\beta_1}}$$

Logit model and estimation

Id	y	x
1	0	1
2	0	4
3	1	5
4	1	6
5	1	9

- Now, take the 3rd observation as an example. Since y=1, it must have been the case that $y^* \geq 0$.
- Thus, the likelihood contribution is:

$$L_3 = P(y_3^* \geq 0) = P(\beta_0 + 5\beta_1 + u_3 \geq 0)$$
$$= P(u_3 \geq -\beta_0 - 5\beta_1) = 1 - \Lambda(-\beta_0 - 5\beta_1)$$
$$= 1 - \frac{1}{1 + e^{-(-\beta_0 - 5\beta_1)}} = 1 - \frac{1}{1 + e^{\beta_0 + 5\beta_1}} = \frac{e^{\beta_0 + 5\beta_1}}{1 + e^{\beta_0 + 5\beta_1}}$$

Logit model and estimation

✓Thus the likelihood function for the data set is given by:

Id	y	x
1	0	1
2	0	4
3	1	5
4	1	6
5	1	9

$$L = \prod_{i=1}^{5} L_i =$$

$$= \frac{1}{1+e^{\beta_0+\beta_1}} \times \frac{1}{1+e^{\beta_0+4\beta_1}} \times \frac{e^{\beta_0+5\beta_1}}{1+e^{\beta_0+5\beta_1}} \times \frac{e^{\beta_0+6\beta_1}}{1+e^{\beta_0+6\beta_1}} \times \frac{e^{\beta_0+9\beta_1}}{1+e^{\beta_0+9\beta_1}}$$

Logit model and estimation

- ✓ Again, we usually maximize the ln(L) instead of the L. Due to invariance, the result is identical.
  ↳ inv. property
- ✓ The values of the parameters that maximize ln(L) are the ML estimators of the (binomial) *logit* model.
- ✓ The MLE is done automatically in Stata and R.

Generalization of the MLE procedure

The likelihood and the log-likelihood function:

$$L = \prod_{i=1}^{n} L_i(\beta) = \prod_{i=1}^{n} (\Lambda(X\beta))^{y_i} \cdot (1 - \Lambda(X\beta))^{1-y_i}$$

$$lnL = \sum_{i=1}^{n} l_i(\beta) = \sum_{i=1}^{n} y_i ln\Lambda(X\beta) + \sum_{i=1}^{n} (1 - y_i) ln(1 - \Lambda(X\beta))$$

The latter is being differentiated:

$\frac{\partial lnL}{\partial x_k} \Rightarrow \hat{\beta}_{k,ML}$ by using numerical methods (iterative procedures).

Marginal or partial effects

the sign before x tells us if the relationship between y and x is positive or negative but the strength is not known

The logit model is also non-linear, therefore the estimated coefficients $\hat{\boldsymbol{\beta}}$ do not reflect the "strength" of the effects of a change in *x* on the probability of occurrence of *y*.

Instead, we calculate the **marginal effects**:

Probability density function of the logistic distr.

- For a continuous x_k: $mfx_k = \frac{\partial p(x)}{\partial x_k} = \lambda(X\beta)\beta_k$.

1. Since $\Lambda(\cdot)$ is strictly increasing, sign of β_k is the same as the sign of $\frac{\partial p(x)}{\partial x_k}$;

2. Relative effects do not depend on *x*: $\frac{\frac{\partial p(x)}{\partial x_k}}{\frac{\partial p(x)}{\partial x_j}} = \frac{\lambda(X\beta)\beta_k}{\lambda(X\beta)\beta_j} = \frac{\beta_k}{\beta_j}$.

Marginal or partial effects

- For a discrete x_k: a change of x_k from c to c+1 results in:

$$\Delta p(x) = \Lambda\big(\beta_0 + \beta_1 x_1 + \cdots + \beta_k (c_k + 1)\big) - \Lambda(\beta_0 + \beta_1 x_1 + \cdots + \beta_k c_k).$$

For a dummy explanatory variable x_k, c = 0:

$$\Delta p(x) = \Lambda(\beta_0 + \beta_1 x_1 + \cdots + \beta_k \cdot 1) - \Lambda(\beta_0 + \beta_1 x_1 + \cdots + \beta_k \cdot 0).$$

In case of the logistic distribution: $\lambda(X\beta)=\Lambda(X\beta)(1-\Lambda(X\beta))$.

Again, most often we evaluate marginal effects at mean values of all explanatory variables, $\lambda(\bar{X}\beta)$, but in general, we can choose any values of our explanatory variables.

Odds, odds ratio and the logit

– The probit model cannot be analyzed like this.

Logit model can also be analyzed in terms of **odds**, i.e. the ratio between the probability of a "positive" outcome (1) and the probabilty of a "negative" outcome (0).

Example for an unspecified course:

- Probability of passing: 3/5
- Probability of failing: 2/5

- Odds of passing (if 1 – pass): $\frac{3/5}{2/5} = \frac{3}{2}$
- Odds of failing (if 1 – fail): $\frac{2/5}{3/5} = \frac{2}{3}$

Odds, odds ratio and the logit

In our case, the **odds Ω** are defined as:

$$\Omega = \frac{P(y=1|x)}{P(y=0|x)} = \frac{P(y=1|x)}{1-P(y=1|x)} =$$

$$= \frac{\frac{e^{X\beta}}{1+e^{X\beta}}}{1-\frac{e^{X\beta}}{1+e^{X\beta}}} = \frac{e^{X\beta}}{1+e^{X\beta}} \cdot \frac{1+e^{X\beta}}{1+e^{X\beta}-e^{X\beta}} = e^{X\beta}$$

The log of odds, also called the **logit**, is then defined as:

$$ln\Omega = \underset{\hookrightarrow ln\, e = 1}{lne^{X\beta}} = X\beta = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

this is useless in practice because odds cannot be observed

The logit model is thus *simply* a linear OLS regression of log of odds on explanatory variables x_j. Unfortunately, we do ***not*** **observe the odds** in practice.

Odds, odds ratio and the logit

We also have the **odds ratio**, *OR*:

odds + one unit of measurement

$$OR_k = \frac{\Omega(X; x_k + 1)}{\Omega(X;\ x_k)} = e^{\beta_k}$$

odds

Interpretation: If x_k increases by 1 unit of measurement, then the odds that $y = 1$ *change* on average, ceteris paribus, by a factor of e^{β_k}.

We thus have three possibilities:

- *OR* = 1: odds unchanged; → this usually doesn't happen in practice
- *OR* > 1: odds increase;
- *OR* < 1: odds decrease.

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

9.1.5 Model Comparison and Interpretation

EFMD EQUIS ACCREDITED

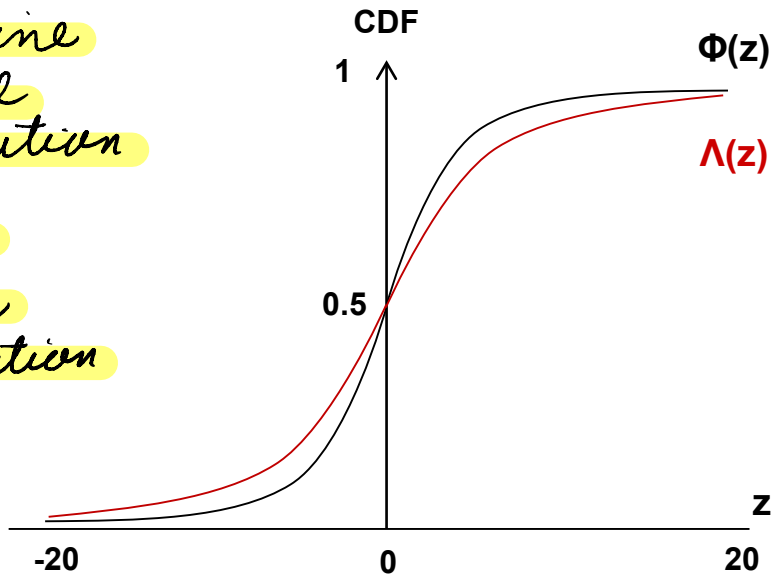
AACSB ACCREDITED

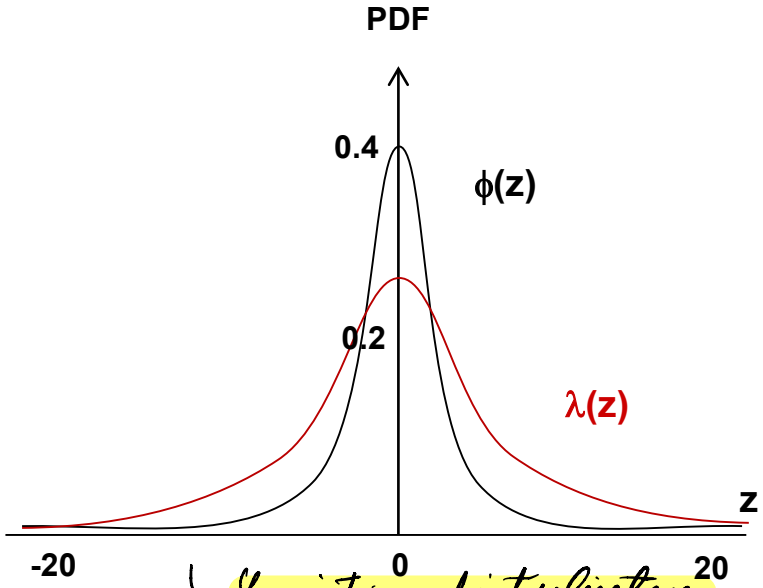
ASSOCIATION AMBA ACCREDITED

Model comparison

black line
↳ normal distribution

red line
↳ logistic distribution





↳ logistic distribution has fatter tails

1. Estimated coefficients differ: $\beta_{logit} \approx 1.7 \cdot \beta_{probit}$.
2. Logistic distribution has "fatter tails", which matters only if the distribution of sample values *y* is extreme (e.g. 95% of 1s).

When we have an extreme distribution (a lot of 1 or 0) we use logistic distribution.

Model interpretation

- ✓ The statistical significance of the regression coefficients can tell you whether the effect of x on the probability of y exists or not.
- ✓ The sign of the regression coefficients can tell you if the probability of y will increase or decrease, when x increases.
- ✓ What if we also want to know "by how much" the probability increases or decreases (the strength of the effect)?
- ✓ This can be done by computing the **partial effects**, also called the **marginal effects**.
- ✓ An alternative approach is to compute the **odds ratios**, done solely based on the *logit model*.

Model interpretation: marginal effects

- ✓ As we already observed, the partial or marginal effect depends on the value of the explanatory variables. Therefore, it is different for every observation in the data.
- ✓ However, we want to know the overall effect of x on the probability of y.
- ✓ For this purpose, we calculate the **partial effect at average**, also called the **marginal effect at average**.

Model interpretation: marginal effects

This is also most commonly done in practice, i.e. the marginal effects are evaluated at mean values of explanatory variables:

/ marginal effect of variable x2

$$\overline{mfx}_k = PEA_k = \phi(\bar{X}\beta)\beta_k \text{ for the probit model;}$$

$$\overline{mfx}_k = PEA_k = \lambda(\bar{X}\beta)\beta_k \text{ for the logit model.}$$

Interpretation of marginal effects at average involves three relativizations: 1) on average, 2) ceteris paribus, and 3) given the mean (average) values of other explanatory variables.

↳ because this is a non-linear model

Model interpretation: marginal effects

Three **most common cases**:

a) Numerical explanatory variable **in levels, x_k**:

If x_k increases by 1 unit of measurement, then the probability that $y = 1$, on average, ceteris paribus, given the means of other explanatory variables, increases/decreases by $100 \cdot \overline{mfx_k}$ percentage points.

↳ probability change is always expressed in percentage points

b) Numerical explanatory variable **in logs, ln x_k**:

If x_k increases by 1 percent, then the probability that $y = 1$, on average, ceteris paribus, given the means of other explanatory variables, increases/decreases by $\overline{mfx_k}$ percentage points.

Model interpretation: marginal effects

c) Explanatory variable is a **dummy variable, *D***:

If $D = 1$, then the probability that $y = 1$, on average, ceteris paribus, given the means of other explanatory variables, increases/decreases by $100 \cdot \overline{mfx}_k$ percentage points, compared to $D = 0$.

Of course, we can evaluate marginal effects at any feasible values of explanatory variables. In that case, we need to use the chosen values of explanatory variables in the calculation and **adjust the third relativization** accordingly.

Alternatively, we could also have calculated the *average partial effect*, which is the mean of the partial effect for each observation.

Model interpretation: odds ratios

Three **most common cases**:

a) Numerical explanatory variable **in levels, x_k**:

$$e^{\ln OR_k \cdot 1} = e^{\beta_k \cdot 1} = e^{\beta_k} = OR_k$$

If x_k increases by 1 unit of measurement, then the odds that $y = 1$, on average, ceteris paribus, increase by $100 \cdot (OR_k - 1)$ percent (if OR > 1) or decrease by $100 \cdot (1 - OR_k)$ percent (if OR < 1).

b) Explanatory variable is a **dummy variable, *D***:

If $D = 1$, then the odds that $y = 1$, on average, ceteris paribus, increase by $100 \cdot (OR_k - 1)$ percent (if OR > 1) or decrease by $100 \cdot (1 - OR_k)$ percent (if OR < 1), compared to $D = 0$.

↳ they either increase or decrease

Model interpretation: odds ratios

c) Numerical explanatory variable **in logs, ln x_k**:

$$e^{\ln OR_k \cdot \ln 1.01} = e^{\beta_k \cdot \ln 1.01}$$

If x_k increases by 1 percent, then the odds that $y = 1$, on average, ceteris paribus, increase by $100 \cdot (e^{\beta_k \cdot \ln 1.01} - 1)$ percent or decrease by $100 \cdot (1 - e^{\beta_k \cdot \ln 1.01})$ percent.

↳ because of logs

Of course, we can evaluate odds ratios for larger changes in explanatory variables. In that case, we need to insert above the chosen change of explanatory variable appropriately.

An example in Stata and R

We have data on several variables concerning the annual holidays:

- *abroad*: dichotomous variable for a person spending holidays abroad: 1 – abroad, 0 – in the home country;
- *log_income*: logarithm of annual family net income per household member;
- *age*: person's age;
- *pet*: dichotomous variable for the presence of pets in the family: 1 – yes, 0 – no.

Estimate the logit model for the *abroad* variable as the dependent variable and all the other variables as explanatory variables.

Calculate the marginal effects at the means of explanatory variables (centroid). Interpret the calculated marginal effects.

An example in Stata

put if the variable is discrete

```
. logit abroad log_income age i.pet

Iteration 0:   log likelihood = -27.525553
Iteration 1:   log likelihood = -17.032774
Iteration 2:   log likelihood = -16.960203
Iteration 3:   log likelihood = -16.959864
Iteration 4:   log likelihood = -16.959864

Logistic regression                             Number of obs     =         40
                                                LR chi2(3)        =      21.13
                                                Prob > chi2       =     0.0001
Log likelihood = -16.959864                     Pseudo R2         =     0.3839

------------------------------------------------------------------------------
      abroad |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  log_income |   4.250643   1.449276     2.93   0.003     1.410114    7.091172
         age |  -.1004235   .0442162    -2.27   0.023    -.1870858   -.0137613
       1.pet |  -2.278437   1.075051    -2.12   0.034    -4.385498    -.171377
       _cons |  -33.26217   12.14418    -2.74   0.006    -57.06433   -9.460013
------------------------------------------------------------------------------
```

$\ln L_{\text{const}}$

$\ln L_{\text{model}}$

$1 - \frac{lnL_{model}}{lnL_{const}}$

There is no multiple determinateion coefficient because we didn't use OLS estimator. Since we used the ML estimator we have Pseudo R^2.

An example in Stata

to obtain marginal effects

```
. margins, dydx(log_income age pet) atmeans

Conditional marginal effects                    Number of obs     =         40
Model VCE    : OIM

Expression   : Pr(abroad), predict()
dy/dx w.r.t. : log_income age 1.pet
at           : log_income      =    9.193383 (mean)
               age             =       47.85 (mean)
               0.pet           =          .7 (mean)
               1.pet           =          .3 (mean)
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
log_income	1.034782	.350304	2.95	0.003	.3481992	1.721366
age	-.0244472	.0108509	-2.25	0.024	-.0457146	-.0031799
1.pet	-.5135038	.1927674	-2.66	0.008	-.891321	-.1356866

Note: dy/dx for factor levels is the discrete change from the base level.

If income increases by 1 percent, then on average, ceteris paribus, given average values of age and pet, the probability of spending holidays abroad increases by 1,03 percentage points.

If age increases by 1 year, then on average, ceteris paribus, given the averages of explanatory variables, the probability of spending holidays abroad decreases by $100 \cdot (-0,0244472)$ percentage points.

If family in the household has a pet, then the probability of spending holidays abroad on average, ceteris paribus, given the averages of explanatory variables decreases by $100 \cdot (-0,5135038)$ percentage points.

An example in R

```
> mod_logit = glm(abroad ~ log_income + age + factor(pet), family=binomial(link="logit"),
 data=holiday)
> summary(mod_logit)

Call:
glm(formula = abroad ~ log_income + age + factor(pet), family = binomial(link = "logit"),
    data = holiday)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7062  -0.6569   0.1760   0.5858   1.9286

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -33.26217   12.14386  -2.739  0.00616 **
log_income    4.25064    1.44923   2.933  0.00336 **
age          -0.10042    0.04422  -2.271  0.02313 *
factor(pet)1 -2.27844    1.07503  -2.119  0.03405 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 55.051  on 39  degrees of freedom
Residual deviance: 33.920  on 36  degrees of freedom
AIC: 41.92

Number of Fisher Scoring iterations: 5
```

An example in R

$1 - \frac{lnL_{model}}{lnL_{const}}$

```
> PseudoR2(mod_logit, which="all")
      McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson
     0.3838502       0.2385307       0.4103844       0.5490215       0.3456715
VeallZimmermann           Efron McKelveyZavoina            Tjur             AIC
     0.5968356       0.4203134       0.6137739       0.4331495      41.9197276
           BIC          logLik          logLik0              G2
    48.6752455     -16.9598638     -27.5255525      21.1313775
> mfx_logit = logitmfx(abroad ~ log_income + age + factor(pet), data=holiday)
> mfx_logit
Call:
logitmfx(formula = abroad ~ log_income + age + factor(pet), data = holiday)

Marginal Effects:
                 dF/dx Std. Err.       z    P>|z|
log_income    1.034783  0.350294  2.9540 0.003136 **
age          -0.024447  0.010851 -2.2531 0.024254 *
factor(pet)1 -0.513504  0.192763 -2.6639 0.007724 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

dF/dx is for discrete change for the following variables:

[1] "factor(pet)1"
```

LR

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

9.2 Multinomial Discrete Choice Models

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Multinomial discrete choice models

- We use multinomial discrete choice models when the individual has to choose among *several discrete alternatives*.
- These alternatives can be **ordered** or **non-ordered**.
- Here we discuss *the latter*, where there is ***no* particular (natural, meaningful) order** among the various options.
- Some **examples**:
  - mode of transport: car, bicycle, bus, train;
  - type of investment: stocks, bonds, options, property;
  - party affiliation: Democrat, independent, Republican;
  - mobile phone brands: Apple, Samsung, Xiaomi, Vivo;
  - occupational field: economist, engineer, lawyer, manager.

Multinomial discrete choice models

- Let us distinguish between the **chooser** and the **choice**:
  - chooser represents an *individual* who has to choose among several alternatives;
  - choice represents the *alternatives* or *options* that face the individual.
- We then distinguish among **three types of models**:
  - multinomial models for chooser-specific or individual-specific data;
  - multinomial models for choice-specific or alternative-specific data;
  - multinomial models for chooser-specific *and* choice-specific data, i.e. mixed data.

Multinomial discrete choice models

- Models for **individual-specific data** include:
  - the multinomial logit model and
  - the multinomial probit model.
- Models for **alternative-specific data** include:
  - the conditional logit model and
  - the alternative-specific multinomial probit model.
- Models for **mixed data** include:
  - the nested logit model and
  - the mixed logit or the random parameters logit model.

We will *only* deal with the **first two groups** of models here.

Independence of irrelevant alternatives

- Some of these models *require* the so-called **independence of irrelevant alternatives** (IIA) assumption.
- **In general**, IIA can be stated as: "If choice A is preferred to choice B out of the choice set {A, B}, then introducing a third alternative X, thus expanding that choice set to {A, B, X}, must *not* make B preferable to A."
- **In case of a multinomial discrete choice model**, IIA implies that adding another alternative or changing the characteristics of a third alternative must *not* affect the relative odds (ratio of choice probabilities) between the two alternatives considered.
- This is ***not* realistic** for many real life applications involving similar (substitute) alternatives.

assumption: introducing a new alternative is not going to alter existing preferences.

↳ in real life this is often not the case

Independence of irrelevant alternatives

- Let us examine the example from McFadden (1974).
- Imagine commuters first facing a decision between two modes of transportation: ***car*** and ***red bus***.
- Suppose that a commumer chooses between these two options with equal probability (0.5), so that the odds ratio equals 1.
- Now add a third mode, ***blue bus***. Assuming that the bus commuters do not care about the colour of the bus (buses are thus perfect substitutes), they are still expected to choose between the bus and the car with equal probability, so the probability of car is *still* 0.5, while the probabilities of each of the two bus types should *decrease* to 0.25.
- However, this violates the IIA; for the *odds ratio* between car and red bus to be *preserved*, the *new probabilities* should be: 0.33 for the car, 0.33 for the red bus, and 0.33 for the blue bus.

Multinomial discrete choice models

- Models that do *not require* the IIA assumption have been developed, but they are *more cumbersome*.
- Models that **require the IIA assumption**:
  - the multinomial logit model;
  - the multinomial probit model and
  - the conditional logit model.
- Models that do ***not* require (loosen) the IIA assumption**:
  - the alternative-specific multinomial probit model;
  - the nested logit model and
  - the mixed logit or the random parameters logit model.

We will *only* deal with the **first group** of models here.

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

9.2.1 Multinomial Models for Individual-specific Data

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Models for individual-specific data

- These models **answer the question**: "How do the *individuals' characteristics* affect their choosing a particular alternative among a set of alternatives?"
- The approach is suitable when the values of **explanatory variables** vary across individuals, but *not* across alternatives (e.g. age or education of the individual).
- Let us start with the **latent variable approach**, in particular with a ***random utility model***, where u_{ij} is the utility for individual i from alternative j, and it is determined by a linear combination of observed characteristics x_i and random disturbance term ε_{ij}:

$$u_{ij} = x_i \beta_j + \varepsilon_{ij}$$

Models for individual-specific data

- An individual chooses alternative k when $u_{ik} > u_{ij}$ for all $j \neq k$. The **probability of choice** for alternative j is:

$$\Pr(y_i = j) = \Pr(u_{ij} > u_{ik} \text{ for all } k \neq j)$$

- The choice is based on the **difference in utilities** between alternatives. So if we assume three alternatives, $J = 3$, and taking one of them as the **base category**, the equations are:

$$u_{i1} - u_{i1} = 0$$
$$u_{i2} - u_{i1} = x_i(\beta_2 - \beta_1) + (\varepsilon_{i2} - \varepsilon_{i1})$$
$$u_{i3} - u_{i1} = x_i(\beta_3 - \beta_1) + (\varepsilon_{i3} - \varepsilon_{i1})$$

logistically distributed

logistically distributed

Models for individual-specific data

- In general, the choice is based on the **difference in choice (latent) variable values** between alternatives.
- The specific form of the model depends on the **distribution of disturbance terms** or, rather, **their differences**.
- In addition, these disturbance terms can be **correlated or not**. If they are *allowed* to be correlated, this can loosen the IIA assumption. If *not*, the IIA assumption is binding.
- Models discussed here go under the names multinomial logit and multinomial probit.
- Both of these models ***require* the IIA assumption**.
- Let us first discuss the **multinomial logit model**.

↳ most widely used

Multinomial logit model

- In essence, this model is like a **set of simultaneous individual binomial logistic regressions**, but **with appropriate weighting**, since the different comparisons between different *pairs of categories* would generally involve different *numbers of observations*.
- We thus generalize the logit model by ***choosing* a base category** and ***setting those coefficients* equal to zero**.
- First, we explain the probability that the *i*-th individual will choose alternative *j*:

$$p_{ij} = P[\textit{individual } i \textit{ chooses alternative } j]$$

- Let us again assume that there are three alternatives, *J* = 3.

Multinomial logit model

base category : β_{11} β_{01}

- For a single explanatory variable, the choice probabilities are:

$$p_{i1} = \frac{1}{1+\exp(\beta_{12}+\beta_{22}x_i)+\exp(\beta_{13}+\beta_{23}x_i)}, j=1$$

$$p_{i2} = \frac{\exp(\beta_{12}+\beta_{22}x_i)}{1+\exp(\beta_{12}+\beta_{22}x_i)+\exp(\beta_{13}+\beta_{23}x_i)}, j=2$$

$$p_{i3} = \frac{\exp(\beta_{13}+\beta_{23}x_i)}{1+\exp(\beta_{12}+\beta_{22}x_i)+\exp(\beta_{13}+\beta_{23}x_i)}, j=3$$

- Again, in this type of models, the explanatory variable(s) is (are) individual specific. That is, it describes the individual, *not* the alternatives facing the individual.
- To distinguish the alternatives, we give them **different parameter values**.

Multinomial logit model

- Let us now turn to **model estimation**.
- Suppose that we observe *three individuals*, who choose alternatives 1, 2, and 3, respectively.
- Assuming that their choices are independent, then the probability of observing this outcome is:

$$P\left[y_{11}=1, y_{22}=1, y_{33}=1\right] = p_{11}\cdot p_{22}\cdot p_{33} =$$

$$= \frac{1}{1+\exp\left(\beta_{12}+\beta_{22}x_1\right)+\exp\left(\beta_{13}+\beta_{23}x_1\right)} \cdot \frac{\exp\left(\beta_{12}+\beta_{22}x_2\right)}{1+\exp\left(\beta_{12}+\beta_{22}x_2\right)+\exp\left(\beta_{13}+\beta_{23}x_2\right)} \cdot$$

$$\cdot \frac{\exp\left(\beta_{13}+\beta_{23}x_3\right)}{1+\exp\left(\beta_{12}+\beta_{22}x_3\right)+\exp\left(\beta_{13}+\beta_{23}x_3\right)} = L\left(\beta_{12},\beta_{22},\beta_{13},\beta_{23}\right)$$

Multinomial logit model

- This is the **likelihood function** of the model.
- The **disturbance terms** of the underlying choice (latent) variable equations are assumed to be uncorrelated; thus the *necessity* for the IIA assumption.
- In addition, they are distributed according to the extreme value distribution, whereas the **difference of two** such **disturbance terms** is then distributed logistically.
- The **maximum likelihood estimator** seeks those values of the parameters that *maximize* the likelihood or, more specifically, the **log-likelihood function**, which is easier to work with mathematically.

Multinomial logit model

- Alternatively, the multinomial logit model could in principle also be **specified *directly*** and **estimated *simultaneously*** as a system of equations defining the logs of relative odds:

$$\ln\left(\frac{p_{i2}}{p_{i1}}\right) = \beta_{12} + \beta_{22} x_i$$

$$\ln\left(\frac{p_{i3}}{p_{i1}}\right) = \beta_{13} + \beta_{23} x_i$$

because the sum of probability is 1

$$p_{i1} = 1 - p_{i2} - p_{i3}$$

- However, this approach suffers from the same *issue* as in the binomial logit model, i.e. we **observe *neither* the probabilities *nor* the odds** in practice.

↳ It is impossible to do this in practice.

Model interpretation

- Similarly to the binomial discrete choice models (binomial logit or probit), the multinomial logit model is also **non-linear**, therefore the estimated coefficients $\hat{\beta}$ do *not* reflect the "*strength*" of the effects of a change in *x* on the probability of *relative* occurrence of a given alternative *j*.
- Yet, regression coefficients *still indicate* the existence (statistical significance) and direction (sign) of the effects of a change in *x* on the probability of *relative* occurrence of a given alternative *j*.
- Nontheless, instead of focusing on particular regression coefficients, we usually calculate and interpret **marginal effects** and **odds ratios**.

Model interpretation: marginal effects

- The **marginal effect** is the effect of a change in *x*, everything else held constant, on the probability that an individual chooses alternative *j* = 1, 2, 3:

$$\left.\frac{\Delta p_{ij}}{\Delta x_i}\right|_{\text{all else constant}} = \frac{\partial p_{ij}}{\partial x_i} = p_{ij}\left[\beta_{2j} - \sum_{m=1}^{3}\beta_{2m}p_{im}\right]$$

- This is marginal change that can and often does *switch sign* at different values of the explanatory variables.
- Alternatively, and somewhat more simply, the *difference in probabilities* can be calculated for two specific values of x_i.

Model interpretation: marginal effects

- If x_a and x_b are the two values of x_i, then the estimated discrete change in probability of choosing alternative 1 [j = 1] when changing from x_a to x_b is:

$$\widetilde{\Delta p_1} = \tilde{p}_{b1} - \tilde{p}_{a1} =$$

$$= \frac{1}{1 + \exp\left(\tilde{\beta}_{12} + \tilde{\beta}_{22} x_b\right) + \exp\left(\tilde{\beta}_{13} + \tilde{\beta}_{23} x_b\right)}$$

$$- \frac{1}{1 + \exp\left(\tilde{\beta}_{12} + \tilde{\beta}_{22} x_a\right) + \exp\left(\tilde{\beta}_{13} + \tilde{\beta}_{23} x_a\right)}$$

- **Interpretation** of these marginal effects is *similar as in the binomial discrete choice models*, but one has to take into account that we now have ***J* of them**, as there are *J* alternatives.

Model interpretation: marginal effects

We compare one of the alternatives to the base.

- This means that for *each* explanatory variable, we will have marginal effects explaining the *change in the probability* of occurence of *each* of the alternatives.
- The interpretation of effects for explanatory variables **in levels**, explanatory variables **in logs** and **discrete** (e.g. dummy) explanatory variables is *preserved* from the binomial discrete choice models.
- In addition, the interpretation of the marginal effects *still* involves the **three relativizations**: 1) on average, 2) ceteris paribus, and 3) given the specific (usually average) values of other explanatory variables (i.e. not subject to interpretation).
- Obviously, as the probabilities always sum up to *one*, the changes in probabilities, i.e. the **marginal effects of a regressor across alternatives sum up to *zero***.

Model interpretation: odds ratios

- In the context of a *multinomial logit model*, the relative **odds ratio**, which is a factor-change coefficient, is also known under the name relative-risk ratio.
- The factor change in the odds of outcome (alternative) *m* versus outcome (alternative) *n* as x_k increases by δ units of measurement, holding other variables constant, equals:

$$\frac{\Omega_{m|n}\left(\mathbf{x}, x_k + \delta\right)}{\Omega_{m|n}\left(\mathbf{x}, x_k\right)} = e^{\beta_{k,m|n}\,\delta}$$

We have several odds ratios => one for each explanatory variable.

- If the amount of change is $\delta = 1$, the relative odds ratio can be interpreted *similarly as in the binomial logit model*.
- Therefore: If x_k increases by 1 unit of measurement, then the odds of outcome *m* versus outcome *n* change on average, ceteris paribus, by a factor of $e^{\beta_{k,m|n}}$.

Model interpretation: odds ratios

- However, *unlike in the binomial logit model*, where we only have *one odds ratio* per explanatory variable (we have *one pair* of outcomes), in the multinomial context one can now consider ***all pairs* of outcomes**.
- Nonetheless, the interpretation of odds ratios for explanatory variables **in levels**, explanatory variables **in logs** and **discrete** (e.g. dummy) explanatory variables is *preserved* from the binomial discrete choice models.
- An interesting feature of the odds ratio is that it **depends *neither*** on the total number of alternatives ***nor*** on the alternatives not examined in the pair.
- There is the *implicit assumption* in logit models that the odds ratio between any pair of alternatives is **independent of irrelevant alternatives** (IIA).

Testing the assumption of IIA

- IIA can be a **fairly strong assumption**, and if it is *violated*, multinomial logit may *not* be a good modeling choice.
- It is especially likely to *fail* if several alternatives are *similar*.
- Two tests of IIA are common in the literature: the **Hausman–McFadden (HM) test** and the **Small–Hsiao (SH) test**.
- The null hypothesis of both tests states that the given odds ratio ***is* independent** of other alternatives. => The assumption is fulfilled (we don't want to reject it)
- Both tests produce *J* variations (each time one alternative is removed and the model is refitted) and thus ***J* results**. They are often **conflicting** (within the run of a test).
- In addition, the SH test requires *randomly dividing the data* into subsamples, thus additional runs produce conflicting results.
- Simulations have shown that both tests have *poor size properties* and are thus **unreliable**. They should be used only with extreme caution.

Multinomial probit model

- Changing the distribution of the vector of **disturbance terms** in the *random utility model* to multivariate normal leads to an *alternative* to the multinomial logit model, called the **multinomial probit model**.
- The disturbance terms of the underlying choice (latent) variable equations are *still* assumed to be uncorrelated; thus the ***necessity* for the IIA assumption**.
- This can be *relaxed* in some versions of the multinomial probit model (e.g. the alternative-specific multinomial probit model), which is *not* the purpose of this section.
- The multinomial logit and multinomial probit model will produce **nearly identical predictions**.
- *Both models* allow calculating the marginal effects and testing hypotheses, but *only the former* allows interpretation in terms of **odds ratios**.

↳ multinomial logit allows to calculate them

Multinomial probit model

If we have extreme distribution we should use the multinomial logit model

- The thicker tails of the *extreme value distribution* used for the multinomial logit model compared with the normal distribution *allow* modelling slightly ***more* extreme behavior**.
- In addition, multinomial probit has ***serious* computational disadvantages**, since it involves calculating multiple ($J - 1$) integrals, a limitation that in the past forced practical applications to only a *few alternatives* (up to 3 or 4).
- Nowadays, quadrature methods can be used to *approximate the integral*, but for large J, this is often *imprecise*, whereas fitting the model still *takes longer*.
- Computational issues make the multinomial probit model *very rare* in practice. *Without further relaxation* (allowing correlation among the disturbance terms and heterogeneity in their distributions) it offers *little value added*.

Multinomial probit is rarely used.

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

9.2.2 Multinomial Models for Alternative-specific Data

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Models for alternative-specific data

- These models **answer the question**: "How do the characteristics or features of various alternatives affect individuals' choice among them?"
- The approach is suitable when the values of **explanatory variables** vary across alternatives (e.g. prices of various products or commuting times by means of transport).
- In this branch of models, alternative-specific variables that *vary over the possible alternatives for each individual* are used to predict which alternative (outcome) is chosen.
- The model discussed here is called the **conditional logit model**. It ***requires* the IIA assumption**.

Independence of irrelevant alternatives

Models for alternative-specific data

- This modelling approach also ***requires* a specific data structure**. In particular, we proceed *from "wide"* (for multinomial logit) *to "long"* form (for conditional logit):

	famid	faminc96	faminc97	faminc98
1.	3	75000	76000	77000
2.	1	40000	40500	41000
3.	2	45000	45400	45800

WIDE

↳ we need to have the "long" form for conditional logit model

	famid	year	faminc
1.	1	96	40000
2.	1	97	40500
3.	1	98	41000
4.	2	96	45000
5.	2	97	45400
6.	2	98	45800
7.	3	96	75000
8.	3	97	76000
9.	3	98	77000

LONG

Conditional logit model

- First, let us consider a model for the probability that individual *i* will choose alternative *j*:

$$p_{ij} = P[\textit{individual } i \textit{ chooses alternative } j]$$

- The conditional logit model specifies these probabilities as:

$$p_{ij} = \frac{\exp\left(\beta_{1j} + \beta_2 x_{ij}\right)}{\exp\left(\beta_{11} + \beta_2 x_{i1}\right) + \exp\left(\beta_{12} + \beta_2 x_{i2}\right) + \exp\left(\beta_{13} + \beta_2 x_{i3}\right)}$$

Intercept for each of them is different, while coefficient is the same (β_2)

- Unlike in the multinomial discrete choice models, there is *only one (common) coefficient*, β_2, relating the effect of the explanatory variable for each alternative to the choice probability p_{ij}, yet the added subscript *j* for an individual *varies across the alternatives*.

Conditional logit model

- We have also included *alternative-specific constants* (ASC, intercept terms). These *cannot all be identified*, and one has to be *set to zero*. We will set $\beta_{13} = 0$.
- Let us now turn to **model estimation**.
- Suppose that we observe *three individuals*, who choose alternatives 1, 2, and 3, respectively.
- Assuming that their choices are independent, then the probability of observing this outcome is:

Conditional logit model

$$P\left(y_{11}=1, y_{22}=1, y_{33}=1\right) = p_{11} \cdot p_{22} \cdot p_{33} =$$

$$= \frac{\exp\left(\beta_{11}+\beta_2 x_{11}\right)}{\exp\left(\beta_{11}+\beta_2 x_{11}\right)+\exp\left(\beta_{12}+\beta_2 x_{12}\right)+\exp\left(\beta_2 x_{13}\right)} \cdot$$

$$\cdot \frac{\exp\left(\beta_{12}+\beta_2 x_{22}\right)}{\exp\left(\beta_{11}+\beta_2 x_{21}\right)+\exp\left(\beta_{12}+\beta_2 x_{22}\right)+\exp\left(\beta_2 x_{23}\right)} \cdot$$

$$\cdot \frac{\exp\left(\beta_2 x_{33}\right)}{\exp\left(\beta_{11}+\beta_2 x_{31}\right)+\exp\left(\beta_{12}+\beta_2 x_{32}\right)+\exp\left(\beta_2 x_{33}\right)} =$$

The intercept is set to zero

$$= L\left(\beta_{12}, \beta_{22}, \beta_2\right)$$

Conditional logit model

- This is also the **likelihood function** of the model.
- The **maximum likelihood estimator** seeks those values of the parameters that *maximize* the likelihood or, more specifically, the **log-likelihood function**, which is easier to work with mathematically.
- We could ***include*** in the regression model **individual-specific explanatory variables** next to the attribute-specific ones. This would lead to a **conditional logit model with individual-specific regressors**.
- If *all* explanatory variables were *individual specific*, then the model becomes identical to the *multinomial logit model*.

Model interpretation

- Similarly to the multinomial logit model, the conditional logit model is also **non-linear**, therefore the same findings apply for the estimated regression coefficients as in the case of the multinomial logit model.
- Even though they *indicate* the existence (statistical significance) and direction (sign) of the effect of a change in explanatory variable related to an alternative on the probability of that alternative being chosen, they do *not* allow for quantitative interpretation ("strength" of the effect) in terms of the probability of that alternative being chosen.
- Instead of focusing on particular regression coefficients, we again calculate and interpret **marginal effects** and **odds ratios**.

Model interpretation: marginal effects

- We shall *not* go into details of the procedures for *calculating* these measures here, instead we will *focus on interpretation*.
- The **marginal effects** are calculated separately *by alternatives*, which means that we obtain ***J* sets of results**.
- *Each set* measures the *change in the probability* of selecting *each alternative* for an increase in the explanatory variable related to *one of the alternatives*.
- Marginal effects are usually calculated *at the alternative-specific means* of explanatory variables, but can in principle be calculated for *any* set of values.
- The interpretation of effects for explanatory variables **in levels**, explanatory variables **in logs** and **discrete** (e.g. dummy) explanatory variables is *preserved* from the binomial discrete choice models.

Model interpretation: marginal effects

- In addition, the interpretation of the marginal effects *still* involves the **three relativizations**: 1) on average, 2) ceteris paribus, and 3) given the specific (usually average) values of other explanatory variables.
- Obviously, as the probabilities always sum up to *one*, the changes in probabilities, i.e. the **marginal effects across alternatives** in each set **sum up to *zero***.
- Should there be *any* *individual-specific explanatory variables* in the model, the corresponding marginal effects are interpreted *just as in the multinomial discrete choice model*.

↳ they are interpreted in the same way.

Model interpretation: odds ratios

- Given that there is *only one* *(common)* *coefficient* relating the effect of an explanatory variable for each alternative to the choice probability, we will have calculated *only one* *(common)* **odds ratio** per explanatory variable, *irrespective of the choice of the base category* (alternative).
- Assuming that the amount of change is one unit of measurement, the relative odds ratio can be interpreted *similarly as in the binomial logit model*.
- Therefore: If x_k related to a given (any) alternative increases by 1 unit of measurement, then the odds of selecting that alternative change on average, ceteris paribus, by a factor of e^{β_k}. "Ceteris paribus" here refers to holding the values of the *given* explanatory variable constant for the *other alternatives*.

Model interpretation: odds ratios

- The interpretation of odds ratios for explanatory variables **in levels**, explanatory variables **in logs** and **discrete** (e.g. dummy) explanatory variables is *preserved* from the binomial discrete choice models.
- The odds ratio for an alternative-specific constant (ASC) indicates the relative likelihood of choosing the *corresponding* alternative compared to the *base* alternative, assuming that the *values of explanatory variables are the same for all alternatives*.
- Should there be *any individual-specific explanatory variables* in the model, the corresponding odds ratios are interpreted *just as in the multinomial discrete choice model*.

Model interpretation: odds ratios

- Again, an interesting feature of the odds ratio is that it **depends *neither*** on the total number of alternatives ***nor*** on the alternatives not currently examined.
- There is the *implicit assumption* in logit models that the odds ratio between any pair of alternatives is **independent of irrelevant alternatives** (IIA).

↳ underlying assumption ⇒ if it is violated then we have an issue and need another model (if the alternatives are substitutes we have a problem)

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

9.3 Ordered Discrete Choice Models

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Ordered discrete choice models

- The choice options in multinomial and conditional logit models have *no natural ordering or arrangement*.
- However, in some cases the choices (discrete response categories) are **ordered** or **ranked** in a specific way.
- Some **examples**:
  - survey scale: strongly disagree, disagree, neutral, agree, strongly agree;
  - assignment of grades: 1, 2, …, 9, 10;
  - bond credit ratings: B, B+, A, A+, A++;
  - work performance ratings: poor, fair, good, very good, outstanding;
  - levels of employment: not in the workforce, working part time, working full time.

we need to use ordered discrete choice model in such case

Ordered discrete choice models

- When modeling these types of outcomes, numerical values are assigned to the outcomes, but the **numerical values** are ordinal, and **reflect *only* the ranking of the outcomes**.
- The **distance** between the values is ***not* meaningful**.
- It is thus *incorrect* to assume that the *distances* between categories are *equal*, as done in the *linear regression model*.
- The **ratio** between any two categories here ***may not* be practically meaningful** either.
- Instead of the *linear* regression model, we employ **ordered discrete choice models**.
- Make sure that the outcome can *indeed* be considered as *ordinal*. E.g., surveys commonly include the category "don't know", which usually does *not correspond to the middle category* in a scale. Such a variable is then *non-ordered*.

it is not ordinal

Latent variable approach

- Let us start with the **latent variable approach**, where y^* is a latent variable and x is a single independent variable:

$$y_i^* = \alpha + \beta x_i + u_i$$

- The latent variable y^* is divided into J ordinal categories:

$$y_i = m \quad \text{if} \quad \mu_{m-1} \leq y_i^* \leq \mu_m \quad \text{for } m = 1 \ldots J$$

- The thresholds or cutpoints μ_1 through μ_{J-1} are estimated. We assume $\mu_0 = -\infty$ and $\mu_J = \infty$.
- Let us consider the problem of choosing what type of college to attend after graduating from high school as an illustration of a choice among unordered alternatives.
- There may be a *natural ordering* to college choice.

Latent variable approach

- When faced with a ranking problem, we develop a *sentiment* about how we feel concerning the alternative choices, and the higher the sentiment, the more likely a higher-ranked alternative will be chosen.
- This is the unobservable, latent variable y^*.
- We might rank the observed *college choices* y as:

if you develop a sentiment then the variable becomes ordinal

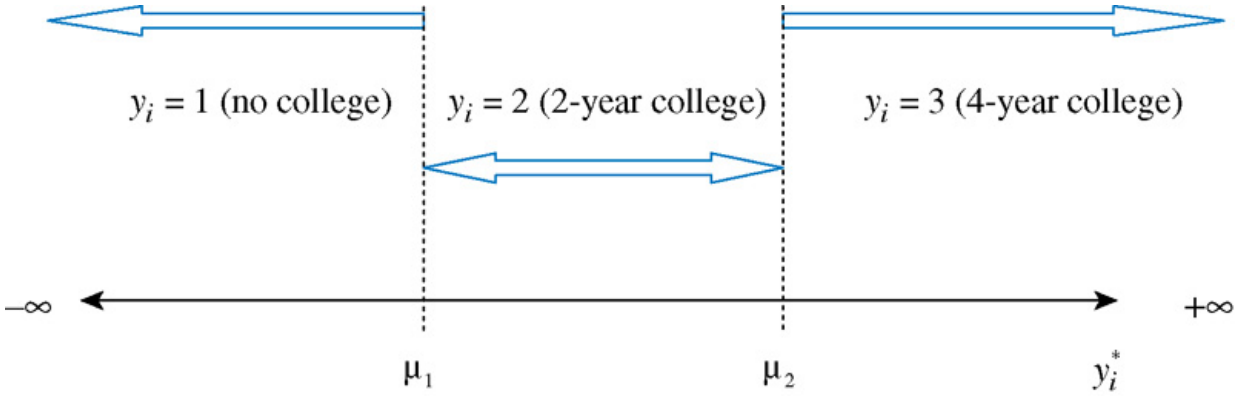
$$y = \begin{cases} 3 & \text{4-year college (the full college experience)} \\ 2 & \text{2-year college (a partial college experience)} \\ 1 & \text{no college} \end{cases}$$

- Again, the linear regression model is *not appropriate* for such data, because in regression we would treat the values y as having some numerical meaning when they *do not*.

sentiment = y^*

Latent variable approach

- When the latent variable *y** crosses a threshold, the observed category *y* changes:



$$y = \begin{cases} 3 \text{ (4-year college)} & \text{if } y_i^* > \mu_2 \\ 2 \text{ (2-year college)} & \text{if } \mu_1 < y_i^* \le \mu_2 \\ 1 \text{ (no college)} & \text{if } y_i^* \le \mu_1 \end{cases}$$

sentiment is greater than μ_2

We need to estimate μ_1 and μ_2 so we can forecast what's going to happen.

Model estimation

- Let us now turn to **model estimation**.
- Suppose that we observe *three individuals*, with the first one not going to college ($y_1 = 1$), the second one attending a two-year college ($y_2 = 2$), and the third one attending a four-year college ($y_3 = 3$).
- Let us further assume that the intercept $\alpha = 0$, so that we can identify (estimate) both thresholds, μ_1 and μ_2.
- For a single explanatory variable, the choice probabilities are:

$$P[y_i = 1] = P[y_i^* \leq \mu_1] = P[\beta x_i + u_i \leq \mu_1] =$$

, u is isolated on the left hand sight

$$= P[u_i \leq \mu_1 - \beta x_i] = \Psi(\mu_1 - \beta x_i)$$

Model estimation

$$P[y_i = 2] = P[\mu_1 < y_i^* \leq \mu_2] = P[\mu_1 < \beta x_i + u_i \leq \mu_2]$$

$$= P[\mu_1 - \beta x_i < u_i \leq \mu_2 - \beta x_i] = \Psi(\mu_2 - \beta x_i) - \Psi(\mu_1 - \beta x_i)$$

$$P[y_i = 3] = P[y_i^* > \mu_2] = P[\beta x_i + u_i > \mu_2] =$$

$$= P[u_i > \mu_2 - \beta x_i] = 1 - \Psi(\mu_2 - \beta x_i)$$

- In the above expressions, $\Psi(\cdot)$ represents a general form of the **cumulative distribution function**.
- Assuming that their choices are independent, then the probability of observing this outcome is:

$$P(y_1 = 1, y_2 = 1, y_3 = 1) =$$
$$= P(y_1 = 1) \cdot P(y_2 = 2) \cdot P(y_3 = 3) = L(\beta, \mu_1, \mu_2)$$

likelyhood with respect to 3 unknowns

Model estimation

- This is the **likelihood function** of the model.
- The **maximum likelihood estimator** seeks those values of the parameters that *maximize* the likelihood or, more specifically, the **log-likelihood function**, which is easier to work with mathematically.
- For maximum likelihood estimation, we need to *assume a form for the cumulative distribution function* of the disturbances.
- *Similarly* to binomial discrete choice models, we have two possibilities: standard normal distribution and logistic distribution.
- In case of the *standard normal* distribution, $\Psi(\cdot) = \Phi(\cdot)$, and we obtain the **ordered probit model**, whereas in case of the *logistic* distribution, $\Psi(\cdot) = \Lambda(\cdot)$, and we obtain the **ordered logit model**.

Model estimation

In finance probit model is more likely to be used

- Most *economists* will use the normality assumption, but many *other social scientists* use the logistic.
- The thicker tails of the *logistic distribution* allow modelling slightly *more extreme behavior*, but in the end, there is ***little difference* between the results**.

Parallel regression assumption

- As observed before, the **slope coefficients** of the explanatory variables are ***the same* in each category**, only their intercepts (tresholds) differ.
- In case of the ordered logit model, the **odds ratio** of the event is independent of the category *j*. The odds ratio is thus assumed to be **constant for all categories**.
- This is called the **parallel regression assumption** or, for the ordered logit model, the proportional odds assumption, Consequently, the *ordered logit model* is also known as the proportional-odds model.
- It *simplifies* the estimation, but it may *not be realistic*.
- The parallel regression assumption can be **tested** by comparing the estimates from $J - 1$ binomial regressions, where the β's are *allowed to differ* across the equations.

Parallel regression assumption

- Several *versions* of the test exist in the literature.
- The null hypothesis of the test states that the *slope coefficients* the binomial regressions are *the same*, which means that the assumption is **not violated**.
- Often times, the hypothesis of parallel regression is *rejected* in practice. However, the test is *sensitive to the number of cases*. Samples with larger numbers of cases are more likely to show a statistically significant test (rejection of the null). Therefore, this test should be used with caution.
- If the assumption *fails*, you may have to consider *other methods*, such as a **multinomial discrete choice model**.

If we have a large sample we will most likely reject the null hypothesis

Model interpretation

- Estimated coefficients (and thresholds) differ between the *ordered logit and probit models*. Estimates from the former are higher than those from the latter by a factor of about 1.7.
- Similarly to the multinomial discrete choice models, the ordered discrete choice models are also **non-linear**.
- Even though they *indicate* the existence (statistical significance) of the effect of a change in explanatory variable on the *choice as such*, interpretation of the direction (sign) and "strength" of the effect on the probability of a *given category being chosen* is *cumbersome*.
- Instead of focusing on particular regression coefficients, we again calculate and interpret **marginal effects** and **odds ratios**.

Model interpretation: marginal effects

- *Similarly as in the binomial discrete choice models*, we distinguish in calculating the **marginal effects** between continuous and discrete explanatory variables.
- For a *continuos explanatory variable*, the marginal effects from an *ordered probit model* in our previous example are:

probability density function

$$\frac{\partial P[y=1]}{\partial x} = -\phi(\mu_1 - \beta x)\cdot\beta$$

$$\frac{\partial P[y=2]}{\partial x} = \left[\phi(\mu_1 - \beta x) - \phi(\mu_2 - \beta x)\right]\cdot\beta$$

$$\frac{\partial P[y=3]}{\partial x} = \phi(\mu_2 - \beta x)\cdot\beta$$

where $\phi(\cdot)$ is the *probability density function* of the standard normal distribution.

Model interpretation: marginal effects

- The *value* of a marginal effect thus depends on the data and the coefficients, whereas the *sign* depends on the density evaluated at two points.
- For a *discrete explanatory variable*, the marginal effects are calculated as differences between evaluated *cumulative distribution functions*.
- *Interpretation* of the marginal effects is *similar as in the binomial discrete choice models*, but one has to take into account that we now have ***J* of them**, as there are *J* choices.
- This means that for *each* explanatory variable, we will have marginal effects explaining the *change in the probability* of a *given category* being chosen.
- Marginal effects are usually calculated *at the averages* of explanatory variables, but can in principle be calculated for *any* set of values.

Model interpretation: marginal effects

- The interpretation of effects for explanatory variables **in levels**, explanatory variables **in logs** and **discrete** (e.g. dummy) explanatory variables is *preserved* from the binomial discrete choice models.
- In addition, the interpretation of the marginal effects *still* involves the **three relativizations**: 1) on average, 2) ceteris paribus, and 3) given the specific (usually average) values of other explanatory variables.
- Obviously, as the probabilities always sum up to *one*, the changes in probabilities, i.e. the **marginal effects across categories sum up to *zero***.
- The marginal effects across a given category change the sign once; the so-called single crossing effect.

Model interpretation: odds ratios

- Based on an estimated *ordered logit model*, we can also calculate the **odds ratios**.
- The factor change in the odds of a lower outcome compared with a higher outcome (for any two *consecutive outcomes*), holding other explanatory variables constant, equals:

$$\frac{\Omega_{\leq m|>n}\left(\mathbf{x}, x_k+\delta\right)}{\Omega_{\leq m|>n}\left(\mathbf{x}, x_k\right)} = e^{-\delta\beta_k} = \frac{1}{e^{\delta\beta_k}}$$

- If the amount of change is $\delta = 1$, the odds ratio can be interpreted *similarly as in the binomial logit model*.
- Therefore: If x_k increases by 1 unit of measurement, then the odds of a *lower* outcome compared with a *higher* outcome change on average, ceteris paribus, by a factor of $e^{-\beta_k}$.

Model interpretation: odds ratios

this isn't that common

- We could also calculate the "reverse" odds ratios, e^{β_k}, i.e. factor changes in the odds of a *higher* outcome compared with a *lower* outcome.
- The interpretation of odds ratios for explanatory variables **in levels**, explanatory variables **in logs** and **discrete** (e.g. dummy) explanatory variables is *preserved* from the binomial discrete choice models.
- As already noted, the value of the **odds ratio does *not* depend on** the value of ***m***, which is why the parallel regression assumption is also known as the **proportional-odds assumption**.

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

9. Discrete Choice Models

Prof. Dr. Miroslav Verbič

miroslav.verbic@ef.uni-lj.si

www.miroslav-verbic.si

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Ljubljana, October 2022