

Multinomial Choice (Basic Models)

Contents

1 Ordered Probit	2
1.1 The Econometric Model	2
1.2 Identification	4
1.3 Interpretation of Parameters	4
1.4 Estimation	5
1.5 Implementation in STATA	5
2 Conditional Logit	7
2.1 The Econometric Model	7
2.2 Identification	8
2.3 Interpretation of Parameters	9
2.4 Estimation	10
2.5 Implementation in STATA	11
3 Multinomial Logit	12
3.1 The Econometric Model	12
3.2 Identification	13
3.3 Interpretation of Parameters	14
3.4 Estimation	15
3.5 Implementation in STATA	15
4 See also ...	16
References	16

1 Ordered Probit

Dependent variables can often only a countable number of values, e.g.

$$y_n \in \{1, 2, \dots, J\}.$$

This applies often to a context where an agent (individual, household, firm, decision maker, ...) chooses from a set of alternatives.

Sometimes the values/categories of such *discrete* variables can be *naturally ordered*, i.e. larger values are assumed to correspond to “higher” outcomes. The ordered probit model is a latent variable model that offers a data generating process for this kind of dependent variables. Some Examples:

- Likert-scale questions in opinion surveys: 1 = “Strongly Disagree”, 2 = “Somewhat Disagree”, 3 = “Undecided”, 4 = “Somewhat Agree”, 5 = “Strongly Agree”.
- Employment status queried as 1 = “unemployed”, 2 = “part time”, 3 = “full time”. (Although often used as example one might question the ‘natural’ order in this case and apply unordered models.)

1.1 The Econometric Model

Consider a latent random variable y_n for individual $n = 1, \dots, N$

$$y_n^* = x_n' \beta + \varepsilon_n \quad \text{with} \quad \varepsilon_i \sim N(0, \sigma^2)$$

that linearly depends on x_n . The error term ε_n is independently and normally distributed with mean 0 and variance σ^2 . The distribution of y_n^* given x_n is therefore also normal: $y_n^* | x_n \sim N(x_n' \beta, \sigma^2)$. The expected value of the latent variable is $E y_n^* = x_n' \beta$.

Observed is only whether individual n 's index lies in a category $j = 1, 2, \dots, J$ which is defined through its unknown lower μ_{j-1} and upper

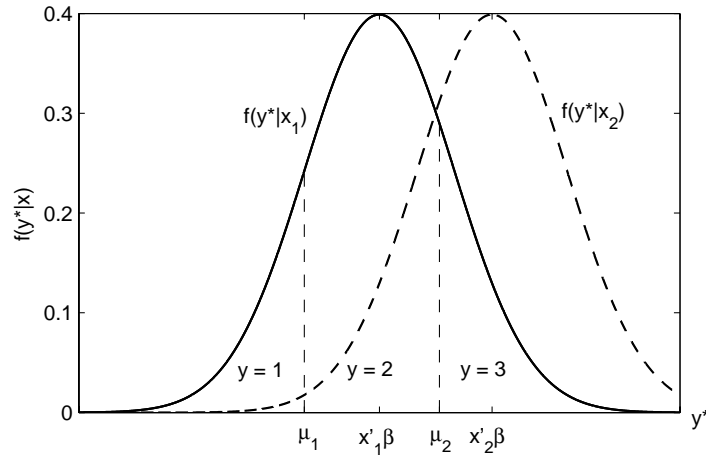


Figure 1: Probabilities in the ordered probit model with 3 alternatives.

bound μ_j , i.e. the observed choice y_n is

$$y_n = \begin{cases} 1 & \text{if } y_n^* \leq \mu_1 \\ 2 & \text{if } \mu_1 < y_n^* \leq \mu_2 \\ 3 & \text{if } \mu_2 < y_n^* \leq \mu_3 \\ \vdots & \\ J & \text{if } \mu_{J-1} < y_n^* \end{cases}$$

The probability that individual n chooses alternative j is easily derived with help of Figure 1:

$$P_{nj} = P(y_n = j|x_n) \begin{cases} \Phi[(\mu_1 - x'_n \beta)/\sigma] & \text{for } j = 1 \\ \Phi[(\mu_2 - x'_n \beta)/\sigma] - \Phi[(\mu_1 - x'_n \beta)/\sigma] & \text{for } j = 2 \\ \Phi[(\mu_3 - x'_n \beta)/\sigma] - \Phi[(\mu_2 - x'_n \beta)/\sigma] & \text{for } j = 3 \\ \vdots & \\ 1 - \Phi[(\mu_{J-1} - x'_n \beta)/\sigma] & \text{for } j = J \end{cases}$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution.

1.2 Identification

The choice probabilities P_{nj} allow only to identify the ratios β/σ and μ/σ but not β , μ and σ individually. Therefore, one usually *assumes* $\sigma = 1$.

Suppose that the index function contains a constant, i.e. $x'_n \beta = \beta_0 + \beta_1 x_1 + \dots + \beta_K x_K$. Then β_0 and μ_1, \dots, μ_{J-1} are not identified as only the differences $(\mu_j - \beta_0)$ appear in the choice probabilities P_{nj} . The model is usually identified by either setting $\mu_1 = 0$ or $\beta_0 = 0$.

1.3 Interpretation of Parameters

[The individual index n is skipped in this section] The sign of the estimated parameters β can directly be interpreted: a positive sign tells whether the answer/choice probabilities shift to higher categories when the independent variable increases. The null hypothesis $\beta_k = 0$ means that the variable x_k , $x = (x_1, \dots, x_k, \dots, x_K)$, has no influence on the choice probabilities. Note, however, that the absolute magnitude of the parameters is meaningless as it is arbitrarily scaled by the assumption $\sigma = 1$. One can therefore e.g. not directly compare parameter estimates for the same variable in different subgroups.

It is often interesting to predict the choice probabilities $P(y = j|x)$ for certain types x and to inspect the marginal effect of an independent variable x_k on the choice probabilities (assuming $\mu_1 = 0$ and $\sigma = 1$)

$$\frac{\partial P(y = 1|x)}{\partial x_k} = -\phi(x' \beta) \beta_k$$

$$\frac{\partial P(y = 2|x)}{\partial x_k} = [\phi(x' \beta) - \phi(\mu_2 - x' \beta)] \beta_k$$

$$\frac{\partial P(y = 3|x)}{\partial x_k} = [\phi(\mu_2 - x' \beta) - \phi(\mu_3 - x' \beta)] \beta_k$$

$$\vdots$$

$$\frac{\partial P(y = J|x)}{\partial x_k} = \phi(\mu_{J-1} - x' \beta) \beta_k$$

Note that the marginal effects can only be reported for specified types x . When β_k is positive, then the probability of choosing the first category $P(y = 1)$ decreases with x_k and the probability of the last category $P(y = J)$ increases. However, the effect on middle categories is ambiguous and depends on x .

1.4 Estimation

The ordered probit model can be estimated using maximum likelihood (ML). The log likelihood function is

$$\log \mathcal{L} = \sum_{n=1}^N \sum_{j=1}^J d_{nj} \log(P_{nj}),$$

where $d_{nj} = 1$ if individual n chooses alternative j and $d_{nj} = 0$ otherwise. The log Likelihood function is numerically maximized subject to $\mu_1 < \mu_2 < \dots < \mu_{J-1}$. The maximum likelihood estimators $\hat{\beta}$ and $\hat{\mu}$ are consistent, asymptotically efficient and normally distributed.

1.5 Implementation in STATA

The Stata command

```
oprobit depvar indepvars
```

estimates the parameter β and the thresholds μ in the ordered probit model. Stata assumes no constant, i.e. $\beta_0 = 0$. **depvar** is a categorical variable which is most favorably but not necessarily coded as 1, ..., J .

The post-estimation command

```
predict p1, p outcome(1)
```

predicts the probability of choosing e.g. the alternative with value $y_n = 1$, in our notation $P(y_n = 1|x_n)$, for all individuals in the sample. You can directly predict the choice probabilities for all alternatives. For $J = 3$ alternatives, the command

```
predict p1 p2 p3, p
```

stores $P(y_n = 1|x_n)$, $P(y_n = 2|x_n)$ and $P(y_n = 3|x_n)$ in the respective new variables **p1**, **p2** and **p3**.

The marginal effects on the probability of choosing the alternative with value 1 is computed by

```
mfx compute, predict(outcome(1))
```

for an individual with mean characteristics \bar{x}_n . The **at** option is used to evaluate further types x_n .

2 Conditional Logit

In most cases, the discrete dependent variables

$$y_n \in \{1, 2, \dots, J\}.$$

have no natural order. This applies often to a context where an agent (individual, household, firm, decision maker, ...) chooses from a *unordered* set of alternatives.

The conditional logit model requires variables that vary across alternatives and possibly across the individuals as well. Some examples:

- Travellers choose among a set of travel modes: “bus”, “train”, “car”, “plane”. There may be a variable “travel time” which is alternative specific and a variable “travel costs” that depends on the travel mode *and* individual income through opportunity costs.
- Car buyers choose among certain types of vehicles: “4-Door Sedans”, “2-Door Coupes”, “Station Wagons”, “Convertibles”, “Sports Cars”, “Mini Vans”, “SUVs”, “Pickup Trucks”, “Vans”.
- Buyers of toilet papers choose among different brands.
- Firms choose from different technologies.

2.1 The Econometric Model

The choice of one out of J unordered alternatives is driven by a latent variable, often interpreted as indirect utility. The indirect utility V_{nj}^* of an individual n choosing alternative $j = 1, \dots, J$ is

$$V_{nj}^* = x'_{nj} \beta + \varepsilon_{nj}.$$

There are J error terms ε_{nj} for any individual n . The exogenous variables $x_{nj} = (x_{nj}^1, x_{nj}^2, x_{nj}^3)$ can be divided into variables that depend only on the individual, x_n^3 , only on the alternative, x_j^2 , or on both x_{nj}^1 .

An individual n chooses alternative j if it offers the highest value of indirect utility. The observed choice y_n of an individual n is therefore

$$y_n = \begin{cases} 1 & \text{if } V_{n1}^* \geq V_{ni}^* \text{ for all } i \\ 2 & \text{if } V_{n2}^* \geq V_{ni}^* \text{ for all } i \\ \vdots & \\ J & \text{if } V_{nJ}^* \geq V_{ni}^* \text{ for all } i \end{cases}$$

Note that this implies that the choice only depends on the *difference* of utility on not on the level.

The conditional logit model assumes that the error terms follow independently and identically an extreme value distribution. The cumulative distribution function is

$$F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}}.$$

This apparently arbitrary specification of the error term has two important features: (1) The difference of two error terms follows a logistic distribution (as in the logit model). (2) The probability that an individual n chooses alternative j is a simple expression (which is not trivial to derive):

$$P_{nj} = P(y_n = j | x_n) = \frac{e^{x'_{nj}\beta}}{\sum_{i=1}^J e^{x'_{ni}\beta}}.$$

The independence of the error term across alternatives is a strong assumption. It implies that an individual’s stochastic, i.e. unobserved, preference for a certain alternative is independent of its stochastic preference for other alternatives. The strong and unpleasant consequences of this assumption are discussed in the literature as *independence of irrelevant alternatives* (IIA).

2.2 Identification

In the conditional logit model, individuals only care about utility differences across alternatives. Factors that influence the level of utility for all

alternatives in the same way can therefore not explain the individual's decision. Individual specific independent variables x_n^3 therefore cancel in the choice probability

$$P_{nj} = \frac{e^{x_{nj}^1 \beta_1 + x_j^2 \beta_2 + x_n^3 \beta_3}}{\sum_{i=1}^J e^{x_{ni}^1 \beta_1 + x_i^2 \beta_2 + x_n^3 \beta_3}} = \frac{e^{x_{nj}^1 \beta_1} e^{x_j^2 \beta_2} e^{x_n^3 \beta_3}}{e^{x_n^3 \beta_3} \sum_{i=1}^J e^{x_{ni}^1 \beta_1} e^{x_i^2 \beta_2}} = \frac{e^{x_{nj}^1 \beta_1} e^{x_j^2 \beta_2}}{\sum_{i=1}^J e^{x_{ni}^1 \beta_1} e^{x_i^2 \beta_2}}$$

and the corresponding β_3 is *not identified*. A *constant* that does neither vary with individuals nor alternatives is of course not identified by the same argument. Individual characteristics x_n^3 start playing a role when they are *interacted* with alternative characteristics x_j^2 .¹

It is often beneficial to include alternative specific constants α_j . These *alternative fixed effect* capture all observed and unobserved characteristics that describe the alternative but are identical across individuals. In this case the coefficient β_2 of the alternative specific variables x_j^2 is *not identified*: any vector q added to $\beta_2^* = \beta_2 + q$ and $\alpha_j^* = \alpha_j - x_j^2 q$ cancels in the choice probabilities P_{nj} . Note that for identification of the fixed effects, one alternative acts as reference and its constant is set to zero.

2.3 Interpretation of Parameters

[*The individual index n is skipped in this section*] In some applications there is a natural interpretation of the latent variable V_j^* . In these situations, the sign of a parameter β_k can be interpreted as the direction of influence of variable x_{jk} , $x_{jk} = (x_{j1}, \dots, x_{jk}, \dots, x_{jK})$ for all j . Note that the absolute magnitude of the parameters is meaningless.

It is sometimes interesting to inspect the marginal effect of an inde-

¹Interacting individual characteristics with alternative fixed effects leads to the multinomial logit model (see the next section).

pendent variable x_{nj} on the choice probabilities:

$$\frac{\partial P(y = j|x)}{\partial x_{jk}} = P_j(1 - P_j)\beta_k$$

$$\frac{\partial P(y = i \neq j|x)}{\partial x_{jk}} = P_j P_i \beta_k$$

Note that the marginal effects depend through P on x and can therefore only be reported for specified types.

It is often most interesting to use the estimated model to predict choice probabilities for specific household types described by x

$$\hat{P}_j = \hat{P}(y = j|x) = \frac{e^{x_j^1 \hat{\beta}}}{\sum_{i=1}^J e^{x_i^1 \hat{\beta}}}$$

The conditional logit model can be used to perform counterfactual policy experiments by varying the values of x_{nj} . Be careful on how policy measures enter the alternative specific x_j^2 and alternative/individual specific characteristics x_{nj}^1 . Note that you cannot inspect alternative specific policy changes when using alternative fixed effects α_j . One can also simulate the effect when new alternatives are added or existing ones are deleted.

2.4 Estimation

The conditional model can be estimated using maximum likelihood (ML). The log likelihood function is

$$\log \mathcal{L} = \sum_{n=1}^N \sum_{j=1}^J d_{nj} \log(P_{nj}),$$

where $d_{nj} = 1$ if individual n chooses alternative j and $d_{nj} = 0$ otherwise. The maximum likelihood estimator $\hat{\beta}$ is consistent, asymptotically efficient and normally distributed.

2.5 Implementation in STATA

Stata requires your data in *long* format when you estimate a conditional logit model: there is a line for any individual n and any alternative j (much like panel data). So the data set contains $N \cdot J$ lines. The dependent variable (`depvar = dnj`) is a dummy variable that indicates if individual n has chosen alternative j or not. The independent variables (`indepvars = xnj`) vary across alternatives and possibly also across individuals. Stata estimates the conditional logit model by the command

```
clogit depvar indepvars, group(groupvar)
```

where the variable (`groupvar = n`) identifies the individual. You can use the post-estimation command

```
predict p, pc1
```

to request predictions of the choice probabilities $P_{nj} = P(y_{nj} = j | x_{nj}) = P(d_{nj} = 1 | x_{nj})$ for all individuals in the sample.

Stata does not provide the marginal effects on the choice probabilities for the conditional logit model.

3 Multinomial Logit

The multinomial logit is used for the same type of choice situations as the conditional logit model:

$$y_n \in \{1, 2, \dots, J\}.$$

where the values of y_n have no natural order.

However, the multinomial logit uses only variables that describe characteristics of the individuals and not of the alternatives. This limits the usefulness of the model for counterfactual predictions. Some examples:

- Travellers choose among a set of travel modes: “bus”, “train”, “car”, “plane”. There are variables that describe the traveller, such as her income. There is no information on the travel modes.
- Car buyers choose among certain types of vehicles: “4-Door Sedans”, “2-Door Coupes”, “Station Wagons”, “Convertibles”, “Sports Cars”, “Mini Vans”, “SUVs”, “Pickup Trucks”, “Vans”. Only information on buyer is used.
- Buyers of toilet papers choose among different brands. Only information on buyer is used.
- Firms choose from different technologies. Only firm information is used.

3.1 The Econometric Model

The multinomial logit model differs from the conditional logit model only in the specification of the deterministic part of the indirect utility V_{nj}^*

$$V_{nj}^* = x_n' \beta_j + \varepsilon_{nj}.$$

The exogenous variables x_n describe only the individual and are identical across alternatives. However the parameter β_j differs across alternatives.

The remaining parts are as in the conditional logit model: The observed choice y_n of an individual n is

$$y_n = \begin{cases} 1 & \text{if } V_{n1}^* \geq V_{ni}^* \text{ for all } i \\ 2 & \text{if } V_{n2}^* \geq V_{ni}^* \text{ for all } i \\ \vdots & \\ J & \text{if } V_{nJ}^* \geq V_{ni}^* \text{ for all } i \end{cases}$$

the error terms follow independently and identically an extreme value distribution

$$F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}},$$

and the probability that an individual n chooses alternative j is

$$P_{nj} = P(y_n = j | x_n) = \frac{e^{x'_n \beta_j}}{\sum_{i=1}^J e^{x'_n \beta_i}}.$$

An interesting feature of the multinomial logit model is that the odds ratio (P_{nj}/P_{ni}) depends log-linearly on x_n

$$\log \left(\frac{P_{nj}}{P_{ni}} \right) = x'_n (\beta_j - \beta_i).$$

3.2 Identification

The parameter vectors β_j , $j = 1, \dots, J$ are not uniquely defined: any vector q added to all vectors $\beta_j^* = \beta_j + q$ cancels in the choice probabilities P_{nj}

$$P_{nj} = \frac{e^{x'_n(\beta_j+q)}}{\sum_{i=1}^J e^{x'_n(\beta_i+q)}} = \frac{e^q e^{x'_n \beta_j}}{e^q \sum_{i=1}^J e^{x'_n \beta_i}} = \frac{e^{x'_n \beta_j}}{\sum_{i=1}^J e^{x'_n \beta_i}}$$

The β_j 's are usually identified by setting the $\beta_i = 0$ for one *reference alternative* i .

3.3 Interpretation of Parameters

[The individual index n is skipped in this section] The parameters of the multinomial logit model are difficult to interpret. Neither the sign (see the identification section above) nor the magnitude of the parameter has an direct intuitive meaning. Hypothesis test have therefore to be very carefully formulated in terms of the estimated parameters.

The marginal effect of an independent variable x_k on the choice probability for alternative j

$$\frac{\partial P(y = j | x)}{\partial x_k} = P_j (\beta_{jk} - \bar{\beta}_k)$$

depends not only on the parameter β_{jk} but also on the mean of all other alternatives $\bar{\beta}_k = 1/J \sum_j \beta_{jk}$.

A potential more direct interpretation of the parameter estimates can be gained by looking at the log of the odds ratio:

$$\frac{\partial \log(P_j/P_i)}{\partial x_k} = \beta_{jk} - \beta_{ik}$$

which reduces to

$$\frac{\partial \log(P_j/P_i)}{\partial x_k} = \beta_{jk}$$

for comparisons with the reference category i . A positive parameter β_{jk} means therefore that the relative probability of choosing j increases relative to the probability of choosing i .

The multinomial logit model can also be used to predict choice probabilities for specific households types x_{nj}

$$\hat{P}_j = \hat{P}(y = j | x) = \frac{e^{x' \hat{\beta}_j}}{\sum_{i=1}^J e^{x' \hat{\beta}_i}}.$$

However one can only inspect changes of individual characteristics on the predicted outcome as all information on the alternatives is enclosed in the estimated alternative specific parameters $\hat{\beta}_j$. Moreover, it is not possible to simulate the addition or deletion of choice alternatives.

3.4 Estimation

The conditional model can be estimated using maximum likelihood (ML). The log likelihood function is

$$\log \mathcal{L} = \sum_{n=1}^N \sum_{j=1}^J d_{nj} \log(P_{nj}),$$

where $d_{nj} = 1$ if individual n chooses alternative j and $d_{nj} = 0$ otherwise. The maximum likelihood estimator $\hat{\beta}$ is consistent, asymptotically efficient and normally distributed.

3.5 Implementation in STATA

The multinomial logit models only uses individual specific characteristics. The data is therefore stored as a usual cross-section dataset: one line per individual. The dependent variable (`depvar = ynj`) is a categorical variable with the individual n 's chosen alternative j . The independent variables (`indepvars = xn`) do not vary across alternatives. Stata estimates the multinomial logit model by the command

```
mlogit depvar indepvars, basecategory(#)
```

where `#` indicates the alternative i for which the parameter $\beta_i = 0$ for identification.

The post-estimation command

```
predict p1, p outcome(1)
```

predicts the probability of choosing the alternative with value $y_n = 1$, in our notation $P(y_n = 1|x_n)$, for all individuals in the sample. You can directly predict the choice probabilities for all alternatives. For 3 alternatives, the command

```
predict p1 p2 p3, p
```

stores $P(y_n = 1|x_n)$, $P(y_n = 2|x_n)$ and $P(y_n = 3|x_n)$ in the respective new variables `p1`, `p2` and `p3`.

The marginal effects on the probability of choosing e.g. the alternative with value 1 is computed by

```
mfx compute, predict(outcome(1))
```

for an individual with mean characteristics \bar{x}_n . The `at` option is used to evaluate further types x_n .

4 See also ...

The independence of irrelevant alternatives (IIA) property of the conditional and the multinomial logit model is in most applications a very unrealistic assumption. The parameter estimators and especially the counterfactual predictions of both models are inconsistent if the IIA does not hold. More flexible models such as *nested logit*, *mixed (kernel) logit* or *multinomial probit* have therefore been proposed. The flexibility of the multinomial probit and the mixed logit model, however, comes at a price: the estimation is numerically very demanding. Moreover, many not yet fully understood practical problems of identification arise.

References

- Train, Kenneth E. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press. Chapter 1 and 2.
- Greene, William H. (2003), *Econometric Analysis*, Prentice Hall. Sections 21.7.1-21.7.3, 21.8.
- Amemiya, Takeshi (1994), *Introduction to Statistics and Econometrics*, Harvard University Press. Section 13.5.2.
- Amemiya, Takeshi (1985), *Advanced Econometrics*, Harvard University Press. Chapter 9.3.1-9.3.4.
- Davidson and MacKinnon (2004), *Econometric Theory and Methods*, Oxford University Press, chapter 11.4.