

3. Multinomial response models

3.1 General model approaches

Multinomial dependent variables in a microeconomic analysis:

These qualitative variables have more than two possible mutually exclusive categories (although binary variables can be considered as special cases of multinomial variables) which are not ordered

Examples for microeconomic analyses with multinomial response models:

- Analysis of the choice of a person among several means of transportation (e.g. car, bus, train)
- Analysis of the employment status of a person (e.g. blue-collar worker, white-collar employee, self-employed person)
- Analysis of the portfolio structure of a household (e.g. no securities, only stocks, only bonds, bonds and stocks)
- Analysis of the choice of a voter among several parties (e.g. CDU/CSU, SPD, Bündnis 90/Die Grünen, Die Linke)
- Analysis of the innovation status of a firm (e.g. no innovations, only product innovations, only process innovations, product and process innovations)
- Analysis of the choice of a car buyer among several energy sources (e.g. gasoline, diesel, hybrid, gas, biofuel, hydrogen, electric)

Utility function of multinomial discrete choice models:

The basis of the microeconomic motivation is that an observation i can choose among J mutually exclusive alternatives of a qualitative variable. As discussed before, the hypothetical (linear) utility function of i for alternative j is as follows:

$$u_{ij} = \beta_j' x_i + \gamma' z_{ij} + \varepsilon_{ij} \quad \text{for } i = 1, \dots, n; j = 1, \dots, J$$

The deterministic component of the utility function comprises the k_1 -dimensional vector $x_i = (x_{i1}, \dots, x_{ik_1})'$ of individual characteristics, the k_2 -dimensional vector $z_{ij} = (z_{ij1}, \dots, z_{ijk_2})'$ of alternative specific attributes, and corresponding parameter vectors $\beta_j = (\beta_{j1}, \dots, \beta_{jk_1})'$ and $\gamma = (\gamma_1, \dots, \gamma_{k_2})'$. The stochastic component of the utility function refers to the error term ε_{ij} that comprises all unobservable factors. The z_{ij} are summarized in the $J \cdot k_2$ -dimensional vector $z_i = (z'_{i1}, \dots, z'_{iJ})'$ and then the x_i and z_i are summarized in the $(k_1 + J \cdot k_2)$ -dimensional vector $X_i = (x'_i, z'_i)'$. The β_j are summarized in the $J \cdot k_1$ -dimensional vector $\beta = (\beta'_1, \dots, \beta'_J)'$.

While the utilities u_{ij} are unobservable, the realizations of the following dummy variables can be observed ($i = 1, \dots, n; j = 1, \dots, J$):

$$y_{ij} = \begin{cases} 1 & \text{if } i \text{ chooses alternative } j \\ 0 & \text{otherwise} \end{cases}$$

According to the stochastic utility maximization hypothesis, observation i chooses category j if the utility of alternative j is the largest of all utilities, i.e. $u_{ij} > u_{ij'}$ ($i = 1, \dots, n; j, j' = 1, \dots, J, j \neq j'$).

Choice probabilities (i.e. probabilities that i chooses j) in multinomial discrete choice models:

$$\begin{aligned}
 p_{ij}(X_i, \beta, \gamma) &= P(y_{ij}=1|X_i, \beta, \gamma) = P(u_{ij} > u_{ij'}, \forall j \neq j'|X_i, \beta, \gamma) \\
 &= P\left(\beta'_j x_i + \gamma' z_{ij} + \varepsilon_{ij} > \beta'_1 x_i + \gamma' z_{i1} + \varepsilon_{i1}; \dots; \right. \\
 &\quad \beta'_j x_i + \gamma' z_{ij} + \varepsilon_{ij} > \beta'_{j-1} x_i + \gamma' z_{i,j-1} + \varepsilon_{i,j-1}; \\
 &\quad \beta'_j x_i + \gamma' z_{ij} + \varepsilon_{ij} > \beta'_{j+1} x_i + \gamma' z_{i,j+1} + \varepsilon_{i,j+1}; \dots; \\
 &\quad \left. \beta'_j x_i + \gamma' z_{ij} + \varepsilon_{ij} > \beta'_J x_i + \gamma' z_{iJ} + \varepsilon_{iJ} \right) \\
 &= P\left(\varepsilon_{i1} - \varepsilon_{ij} < (\beta'_j x_i + \gamma' z_{ij}) - (\beta'_1 x_i + \gamma' z_{i1}); \dots; \right. \\
 &\quad \varepsilon_{i,j-1} - \varepsilon_{ij} < (\beta'_j x_i + \gamma' z_{ij}) - (\beta'_{j-1} x_i + \gamma' z_{i,j-1}); \\
 &\quad \varepsilon_{i,j+1} - \varepsilon_{ij} < (\beta'_j x_i + \gamma' z_{ij}) - (\beta'_{j+1} x_i + \gamma' z_{i,j+1}); \dots; \\
 &\quad \left. \varepsilon_{iJ} - \varepsilon_{ij} < (\beta'_j x_i + \gamma' z_{ij}) - (\beta'_J x_i + \gamma' z_{iJ}) \right) \\
 &= P\left(\varepsilon_{ij'} - \varepsilon_{ij} < (\beta'_j x_i + \gamma' z_{ij}) - (\beta'_{j'} x_i + \gamma' z_{ij'}); \forall j \neq j'\right)
 \end{aligned}$$

These choice probabilities are the basis for the discrete choice analysis. Different distribution assumptions for the stochastic component ε_{ij} lead to different choice probabilities and thus to different multinomial discrete choice models. The special case of $J = 2$ leads to binary discrete choice models.

ML estimation in multinomial discrete choice models:

In the following, the J-dimensional vector $y_i = (y_{i1}, \dots, y_{iJ})'$ comprises the observable dependent variables as discussed above and X_i comprises all explanatory variables. Furthermore, all (free) parameters (particularly in β and γ , but possibly also variance covariance parameters, see later) are summarized in the vector θ . In the case of multinomial discrete choice models, the y_i are multinomially distributed with the parameters 1 and the choice probabilities $p_{ij}(X_i, \beta, \gamma)$. Based on a random sample (X_i, y_i) for $i = 1, \dots, n$ observations, the likelihood and log-likelihood functions are:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n f_i(y_i; X_i, \theta) = \prod_{i=1}^n p_{i1}(X_i, \theta)^{y_{i1}} p_{i2}(X_i, \theta)^{y_{i2}} \cdots p_{iJ}(X_i, \theta)^{y_{iJ}} \\ &= \prod_{i=1}^n \prod_{j=1}^J p_{ij}(X_i, \theta)^{y_{ij}} \end{aligned}$$

$$\log L(\theta) = \sum_{i=1}^n \sum_{j=1}^J y_{ij} \log p_{ij}(X_i, \theta)$$

The ML estimator solves the first-order conditions for maximizing the log-likelihood function. Thus by equalizing the score with zero it follows:

$$\hat{\theta} = \underset{\theta}{\text{arg solves}} \left[\sum_{i=1}^n s_i(\theta) = \sum_{i=1}^n \sum_{j=1}^J y_{ij} \frac{\partial \log p_{ij}(X_i, \theta)}{\partial \theta} = 0 \right]$$

Fundamental distribution assumption for multinomial logit models:

The error terms ε_{ij} are independently and identically standard extreme value distributed over all categories $j = 1, \dots, J$. With this assumption, a single difference of two ε_{ij} has a standard logistic distribution. In the special case of $J = 2$ the multinomial logit model falls back to the binary logit model.

Choice probabilities in general multinomial logit models ($i = 1, \dots, n; j = 1, \dots, J$):

$$p_{ij}(X_i, \beta, \gamma) = P(y_{ij}=1|X_i, \beta, \gamma) = \frac{e^{\beta_j'x_i + \gamma'z_{ij}}}{\sum_{m=1}^J e^{\beta_m'x_i + \gamma'z_{im}}}$$

As required, these values vary between zero and one and add up to one over all j .

However, these choice probabilities comprise too many parameters in β and thus are not identified since any constant can be added to each of the parameter vectors β_1, \dots, β_J without changing the probabilities, i.e. only the differences between β_1, \dots, β_J are relevant. Therefore, one of these vectors has to be parameterized. Common approaches are to set the parameter vector for alternative 1 or for alternative J to zero, i.e. $\beta_1 = 0$ or $\beta_J = 0$. In the following, we consider the second approach.

On the basis of this normalization $\beta_J = 0$, the category J is the base category (or baseline) and provides the reference point for all other alternatives. This has to be considered for the interpretation of the estimation results (see later). If the numerator and denominator of the choice probabilities are divided by $e^{\beta_J'x_i + \gamma'z_{iJ}} = e^{0 + \gamma'z_{iJ}} = e^{\gamma'z_{iJ}}$, it follows:

$$p_{ij}(X_i, \beta, \gamma) = \frac{e^{\beta_j'x_i + \gamma'(z_{ij} - z_{iJ})}}{1 + \sum_{m=1}^{J-1} e^{\beta_m'x_i + \gamma'(z_{im} - z_{iJ})}} \quad \text{for } j = 1, \dots, J-1$$

$$p_{iJ}(X_i, \beta, \gamma) = \frac{1}{1 + \sum_{m=1}^{J-1} e^{\beta_m'x_i + \gamma'(z_{im} - z_{iJ})}}$$

→ These choice probabilities refer to the most flexible multinomial logit model approach which includes both individual characteristics and alternative specific attributes as explanatory variables. In many empirical studies, however, only one class of explanatory variables is examined. While the term “multinomial logit model” is not consistently used, it often refers to model approaches that exclusively include individual characteristics. Approaches with only alternative specific attributes as explanatory variables are often called “conditional logit models”.

3.2 (Pure) multinomial logit models

Choice probabilities in (pure) multinomial logit models ($i = 1, \dots, n$; $j = 1, \dots, J$):

$$p_{ij}(x_i, \beta) = \frac{e^{\beta'_j x_i}}{\sum_{m=1}^J e^{\beta'_m x_i}}$$

Based on the aforementioned parameterization $\beta_j = 0$, the choice probabilities in such approaches can be alternatively written as follows:

$$p_{ij}(x_i, \beta) = \frac{e^{\beta'_j x_i}}{1 + \sum_{m=1}^{J-1} e^{\beta'_m x_i}} \quad \text{for } j = 1, \dots, J-1$$

$$p_{iJ}(x_i, \beta) = \frac{1}{1 + \sum_{m=1}^{J-1} e^{\beta'_m x_i}}$$

The inclusion of the ML estimator $\hat{\beta}$ into the choice probabilities leads to the corresponding estimator $\hat{p}_{ij}(x_i, \hat{\beta})$ of the choice probabilities for all categories $j = 1, \dots, J$. According to these formulas, the (estimators of) choice probabilities for alternative j imply that they do not only depend on the (estimator of the) parameter vector β_j , but on all other (estimators of) parameter vectors.

In line with binary logit models (which are a special case of these multinomial logit models) and binary probit models, the parameter estimators furthermore cannot be interpreted like the estimators of the effect of the respective explanatory variable. Instead, it follows for the estimator of the (partial) marginal probability effect of a (continuous) individual characteristic x_{ih} as explanatory variable in (pure) multinomial logit models ($i = 1, \dots, n$):

$$\frac{\partial \hat{p}_{ij}(x_i, \hat{\beta})}{\partial x_{ih}} = \hat{p}_{ij}(x_i, \hat{\beta}) \left[\hat{\beta}_{jh} - \sum_{m=1}^{J-1} \hat{p}_{im}(x_i, \hat{\beta}) \hat{\beta}_{mh} \right] \quad \text{for } j = 1, \dots, J-1$$

$$\frac{\partial \hat{p}_{iJ}(x_i, \hat{\beta})}{\partial x_{ih}} = -\hat{p}_{iJ}(x_i, \hat{\beta}) \sum_{m=1}^{J-1} \hat{p}_{im}(x_i, \hat{\beta}) \hat{\beta}_{mh}$$

Interpretation:

- This formula refers to the estimator of the effect of a small infinitesimal increase of x_{ih} on the change of the probability to choose alternative j
- As aforementioned, this estimator of the marginal probability effect does not only depend on the ML estimator $\hat{\beta}_{jh}$ for j , but also on the estimators of the choice probabilities and thus the parameters for all other categories. Furthermore, it varies with different values of all individual characteristics.
- In contrast to the case of binary logit models, $\hat{\beta}_{jh}$ not even indicates the direction of the estimator of marginal probability effects, i.e. a positive (negative) $\hat{\beta}_{jh}$ does not necessarily lead to positive (negative) estimators

Based on y_1, \dots, y_n and x_1, \dots, x_n , it follows for the estimator of the average (partial) marginal probability effect ($AMPE_{hj}$) of the individual characteristic x_{ih} across all i in (pure) multinomial logit models:

$$\hat{AMPE}_{hj} = \frac{1}{n} \sum_{i=1}^n \hat{p}_{ij}(x_i, \hat{\beta}) \left[\hat{\beta}_{jh} - \sum_{m=1}^{J-1} \hat{p}_{im}(x_i, \hat{\beta}) \hat{\beta}_{mh} \right] \quad \text{for } j = 1, \dots, J-1$$

$$\hat{AMPE}_{hj} = \frac{1}{n} \sum_{i=1}^n \left[-\hat{p}_{ij}(x_i, \hat{\beta}) \sum_{m=1}^{J-1} \hat{p}_{im}(x_i, \hat{\beta}) \hat{\beta}_{mh} \right]$$

The (partial) marginal probability effect at the means of the individual characteristics across $i = 1, \dots, n$ can be correspondingly estimated.

For discrete or qualitative individual characteristics as explanatory variables and larger changes of (continuous) individual characteristics the estimator of marginal probability effects can again lead to very inaccurate results. The estimator of a discrete change of the choice probabilities $p_{ij}(x_i, \beta)$ due to a discrete change Δx_{ih} in (pure) multinomial logit models is as follows (for $j = 1, \dots, J-1$):

$$\begin{aligned} \Delta \hat{p}_{ij}(x_i, \hat{\beta}) &= \Delta P(y_{ij}=1|x_i, \hat{\beta}) = P(y_{ij}=1|x_i + \Delta x_{ih}, \hat{\beta}) - P(y_{ij}=1|x_i, \hat{\beta}) \\ &= \frac{e^{\hat{\beta}'_j x_i + \hat{\beta}_{jh} \Delta x_{ih}}}{1 + \sum_{m=1}^{J-1} e^{\hat{\beta}'_m x_i + \hat{\beta}_{mh} \Delta x_{ih}}} - \frac{e^{\hat{\beta}'_j x_i}}{1 + \sum_{m=1}^{J-1} e^{\hat{\beta}'_m x_i}} \end{aligned}$$

Since the sum over the estimated choice probabilities for all alternatives must be equal to one, the change of one estimator of probabilities is determined by the $J-1$ other changes so that it follows for the estimator of a discrete change of the choice probability $p_{ij}(x_i, \beta)$ due to Δx_{ih} :

$$\Delta \hat{p}_{ij}(x_i, \hat{\beta}) = - \sum_{j=1}^{J-1} \Delta \hat{p}_{ij}(x_i, \hat{\beta})$$

Remarks:

- As in the case of estimated marginal probability effects, the sign of the estimated change $\Delta \hat{p}_{ij}(x_i, \hat{\beta})$ for all $j = 1, \dots, J$ due to a discrete change Δx_{ih} of the individual characteristic x_{ih} need not coincide with the sign of the corresponding ML estimator $\hat{\beta}_{jh}$ for j . If e.g. $\hat{\beta}_{jh}$ is positive, the numerator of the first term in $\Delta \hat{p}_{ij}(x_i, \hat{\beta})$ increases with increasing x_{ih} . However, it is possible that the denominator increases even more due to the values $\hat{\beta}_{mh}$ ($\forall m \neq j$).
- As in the case of estimated marginal probability effects, the estimated changes $\Delta \hat{p}_{ij}(x_i, \hat{\beta})$ vary with different values not only of x_{ih} but also with different values of all other individual characteristics and thus across different observations
- On this basis, average discrete changes of $p_{ij}(x_i, \beta)$ ($j = 1, \dots, J$) across all i and corresponding discrete changes of $p_{ij}(x_i, \beta)$ at the means of the individual characteristics across $i = 1, \dots, n$ can be estimated

While the ML estimator $\hat{\beta}_{jh}$ neither indicates the extent nor the direction of the effect of an individual characteristic x_{ih} on the estimator $\hat{p}_{ij}(x_i, \hat{\beta})$ of the choice probability for alternative j , it nevertheless has an important informative value. This can be recognized by dividing the estimator $\hat{p}_{ij}(x_i, \hat{\beta})$ of the choice probability for j and the corresponding estimator $\hat{p}_{iJ}(x_i, \hat{\beta})$ for the base category J . For the so-called odds it follows for $j = 1, \dots, J-1$:

$$\frac{\hat{p}_{ij}(x_i, \hat{\beta})}{\hat{p}_{iJ}(x_i, \hat{\beta})} = \frac{e^{\hat{\beta}'_j x_i}}{1 + \sum_{m=1}^{J-1} e^{\hat{\beta}'_m x_i}} = e^{\hat{\beta}'_j x_i} = e^{\hat{\beta}'_{j1} x_{i1} + \dots + \hat{\beta}'_{jk_1} x_{ik_1}}$$

Interpretation:

This formula shows that although the ML estimator $\hat{\beta}_{jh}$ does not indicate the effect of x_{ih} on the estimator $\hat{p}_{ij}(x_i, \hat{\beta})$ of the choice probability for j alone, it indicates the direction of the effect on $\hat{p}_{ij}(x_i, \hat{\beta})$ relative to the estimator $\hat{p}_{iJ}(x_i, \hat{\beta})$ of the choice probability for the base category J . If $\hat{\beta}_{jh}$ is positive (negative), an increase of x_{ih} increases (decreases) the odds, i.e. $\hat{p}_{ij}(x_i, \hat{\beta})$ relative to $\hat{p}_{iJ}(x_i, \hat{\beta})$.

The previous analysis of the estimation of the probability effects relative to the base category can be extended to the discussion of the odds for two arbitrary alternatives j and r . It follows ($\forall r \neq j$):

$$\frac{\hat{p}_{ij}(x_i, \hat{\beta})}{\hat{p}_{ir}(x_i, \hat{\beta})} = \frac{\frac{e^{\hat{\beta}'_j x_i}}{1 + \sum_{m=1}^{J-1} e^{\hat{\beta}'_m x_i}}}{\frac{e^{\hat{\beta}'_r x_i}}{1 + \sum_{m=1}^{J-1} e^{\hat{\beta}'_m x_i}}} = \frac{e^{\hat{\beta}'_j x_i}}{e^{\hat{\beta}'_r x_i}} = e^{(\hat{\beta}'_j - \hat{\beta}'_r) x_i} = e^{(\hat{\beta}_{j1} - \hat{\beta}_{r1}) x_{i1} + \dots + (\hat{\beta}_{jk_1} - \hat{\beta}_{rk_1}) x_{ik_1}}$$

Interpretation:

This formula implies that the difference between the two ML estimators $\hat{\beta}_{jh}$ and $\hat{\beta}_{rh}$ indicates the direction of the effect of x_{ih} on the estimator $\hat{p}_{ij}(x_i, \hat{\beta})$ of the choice probability for category j relative to the estimator $\hat{p}_{ir}(x_i, \hat{\beta})$ of the choice probability for category r . If $\hat{\beta}_{jh}$ is greater (less) than $\hat{\beta}_{rh}$, an increase of x_{ih} increases (decreases) $\hat{p}_{ij}(x_i, \hat{\beta})$ relative to $\hat{p}_{ir}(x_i, \hat{\beta})$.

Example: Determinants of secondary school choice (I)

By using a (pure) multinomial logit model, the effect of the following individual characteristics on the choice of 675 pupils in Germany between the three secondary school types Hauptschule, Realschule, and Gymnasium is analyzed:

- Years of education of the mother (motheduc) as mainly interesting explanatory variable
- Dummy variable for labor force participation of the mother (mothlnlf) that takes the value one if the mother is employed
- Logarithm of household income (loghhincome)
- Logarithm of household size (loghhsize)
- Rank by age among the siblings (birthorder)
- Year dummies for 1995-2002

The three alternatives of the multinomial dependent variable secondary school (schooltype) take the values one for Hauptschule, two for Realschule, and three for Gymnasium, whereby Hauptschule is chosen as base category. As a consequence, two vectors of parameters for the alternatives Realschule and Gymnasium are estimated. The ML estimation of the multinomial logit model with STATA leads to the following results:

Example: Determinants of secondary school choice (II)

```
mlogit schooltype motheduc mothinlf loghhincome loghhsizе birthorder year1995 year1996
year1997 year1998 year1999 year2000 year2001 year2002, base(1)
```

```
Multinomial logistic regression          Number of obs   =          675
                                          LR chi2(26)     =          221.20
                                          Prob > chi2     =           0.0000
Log likelihood = -622.24169              Pseudo R2       =           0.1509
```

	schooltype	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1		(base outcome)					
2							
	motheduc	.2987624	.0789985	3.78	0.000	.1439282	.4535967
	mothinlf	-.3910109	.2284116	-1.71	0.087	-.8386895	.0566677
	loghhincome	.4073928	.2254481	1.81	0.071	-.0344774	.8492631
	loghhsizе	-1.145075	.4452722	-2.57	0.010	-2.017792	-.2723571
	birthorder	-.1229556	.1255023	-0.98	0.327	-.3689357	.1230245
	year1995	.0767306	.4236393	0.18	0.856	-.753587	.9070483
	year1996	.2193122	.4457021	0.49	0.623	-.6542478	1.092872
	year1997	-.1001198	.4370522	-0.23	0.819	-.9567264	.7564868
	year1998	.4394545	.4872894	0.90	0.367	-.5156151	1.394524
	year1999	.5753325	.4617021	1.25	0.213	-.3295869	1.480252
	year2000	.1609341	.4550522	0.35	0.724	-.7309519	1.05282
	year2001	.4590218	.4554232	1.01	0.314	-.4335913	1.351635
	year2002	.1428137	.4488494	0.32	0.750	-.7369149	1.022542
	_cons	-5.795185	2.312289	-2.51	0.012	-10.32719	-1.263182

Example: Determinants of secondary school choice (III)

3							
motheduc		.6554335	.0811063	8.08	0.000	.496468	.814399
mothlnlf		-.3775209	.2353811	-1.60	0.109	-.8388595	.0838177
loghhincome		1.710194	.2832594	6.04	0.000	1.155016	2.265372
loghhsiz		-1.471701	.4843607	-3.04	0.002	-2.42103	-.5223712
birthorder		-.2736975	.1359262	-2.01	0.044	-.540108	-.007287
year1995		.0761265	.4281856	0.18	0.859	-.7631019	.915355
year1996		.1559678	.4471054	0.35	0.727	-.7203426	1.032278
year1997		-.6671805	.4582662	-1.46	0.145	-1.565366	.2310047
year1998		.1200219	.4989005	0.24	0.810	-.8578052	1.097849
year1999		-.3979308	.4970546	-0.80	0.423	-1.37214	.5762783
year2000		-.0598782	.4663055	-0.13	0.898	-.9738202	.8540638
year2001		.0855021	.4662402	0.18	0.854	-.8283119	.9993161
year2002		-.3087189	.4494414	-0.69	0.492	-1.189608	.5721701
_cons		-23.05832	2.962141	-7.78	0.000	-28.86401	-17.25263

As already discussed in the analysis of binary logit and probit models, the presentation of estimation results in empirical studies particularly comprises the parameter estimates, the z statistics or estimated standard deviations of the estimated parameters, and some information about the significance of the rejection of the null hypothesis that the parameter is zero. An exemplary table can have the following form:

Example: Determinants of secondary school choice (IV)

ML estimates (z statistics), dependent variable: school type, base category: Hauptschule		
Explanatory variables	Realschule	Gymnasium
motheduc	0.299*** (3.78)	0.655*** (8.08)
mothlnlf	-0.391* (-1.71)	-0.378 (-1.60)
loghhincome	0.407* (1.81)	1.710 *** (6.04)
loghhszise	-1.145** (-2.57)	-1.472*** (-3.04)
birthorder	-0.123 (-0.98)	-0.274** (-2.01)
constant	-5.795** (-2.51)	-23.058*** (-7.78)
Year dummies 1995-2002	Yes	
Maximum value of log-likelihood function	-622.24	
Likelihood ratio test statistic (all parameters)	221.20***	

Note: *** (**, *) means that the appropriate parameter is different from zero or that the underlying null hypothesis is rejected at the 1% (5%, 10%) significance level, n = 675

Example: Determinants of secondary school choice (V)

Interpretation:

- The value of 221.20 for the likelihood ratio test statistic implies that the null hypothesis that all 26 parameters are zero (which would imply that no explanatory variable has an effect on the choice of Realschule or Gymnasium relative to Hauptschule) can be rejected at any common significance level
- The parameter estimates for motheduc are positive for both alternatives Realschule and Gymnasium and highly significantly different from zero due to the z statistics of 3.78 for Realschule and 8.08 for Gymnasium
- These parameter estimates therefore imply that the years of education of the mother have a strong significantly positive effect on the (probability of the) choice of Realschule compared with Hauptschule and additionally on the (probability of the) choice of Gymnasium compared with Hauptschule
- The negative value of the difference $0.299 - 0.655 = -0.356$ of the parameter estimates for motheduc for Realschule and Gymnasium implies that the years of education of the mother have a negative effect on the choice of Realschule relative to Gymnasium or conversely a positive effect on the choice of Gymnasium relative to Realschule. The significance of these effects has to be analyzed by choosing Realschule or Gymnasium as base category.

Example: Determinants of secondary school choice (VI)

Wald and likelihood ratio tests:

As an example, the null hypothesis that motheduc has no effect on the secondary school choice, i.e. that the two corresponding parameters are zero, is tested. The command for the Wald test in STATA is:

```
test motheduc

( 1)  [Hauptschule]o.motheduc = 0
( 2)  [Realschule]motheduc = 0
( 3)  [Gymnasium]motheduc = 0
      Constraint 1 dropped

      chi2( 2) =    73.70
      Prob > chi2 =    0.0000
```

With respect to the application of the likelihood ratio test, the STATA command “estimates store unrestr” after the unrestricted ML estimation and the command “estimates store restr” after the restricted ML estimation are necessary (the choice of the names is again arbitrary). The command for the likelihood ratio test in STATA is then:

```
lrtest unrestr restr

Likelihood-ratio test                                LR chi2(2) =    107.99
(Assumption: restr nested in unrestr)              Prob > chi2 =    0.0000
```

Example: Determinants of secondary school choice (VII)

The estimation of the average marginal probability effect of motheduc across all 675 pupils on the choice of Gymnasium with STATA leads to the following results:

```
margins, dydx(motheduc) predict(outcome(3))
```

```
Average marginal effects          Number of obs   =          675
```

```
Model VCE      : OIM
```

```
Expression     : Pr(schooltype==3), predict(outcome(3))
```

```
dy/dx w.r.t.  : motheduc
```

		Delta-method				[95% Conf. Interval]	
	dy/dx	Std. Err.	z	P> z			
motheduc	.0886491	.0084999	10.43	0.000	.0719897		.1053085

This value of 0.0886 means that an increase of the years of education of the mother by one (unit) leads to an approximately estimated increase of the choice probability for Gymnasium by 8.86 percentage points. The corresponding values for Hauptschule and Realschule are -0.0797 and -0.0089. These values differ from the estimates of the marginal probability effect at the means of the individual characteristics across all 675 pupils. For the effects on the choice of Gymnasium the estimation with STATA leads to the following results;₁₉

Example: Determinants of secondary school choice (VIII)

```
margins, dydx(motheduc) atmeans predict(outcome(3))
```

```
Conditional marginal effects                               Number of obs =           675
```

```
Model VCE      : OIM
```

```
Expression      : Pr(schooltype==3), predict(outcome(3))
```

```
dy/dx w.r.t.    : motheduc
```

```
at              : motheduc           =    11.44296 (mean)
                  mothlnlf          =     .5525926 (mean)
                  loghhincome       =    11.05839 (mean)
                  loghhsz           =     1.412881 (mean)
                  birthorder        =         1.76 (mean)
                  year1995          =     .1377778 (mean)
                  year1996          =         .12 (mean)
                  year1997          =     .1111111 (mean)
                  year1998          =     .0888889 (mean)
                  year1999          =     .1007407 (mean)
                  year2000          =     .1037037 (mean)
                  year2001          =     .1185185 (mean)
                  year2002          =     .117037 (mean)
```

		Delta-method				
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
motheduc	.1170922	.0140003	8.36	0.000	.0896521	.1445323

Example: Determinants of secondary school choice (IX)

The estimation of the average probabilities of the choice of Gymnasium across all 675 pupils for the minimum and maximum values of motheduc = 7 and motheduc = 18 years with STATA leads to the following results:

```
margins, at(motheduc=7) predict(outcome(3))
```

```
Predictive margins                                Number of obs   =           675
Model VCE      : OIM
Expression     : Pr(schooltype==Gymnasium), predict(outcome(3))
at             : motheduc           =           7
```

	Margin	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
---+---						
_cons	.0747206	.0196424	3.80	0.000	.0362222	.1132189

```
margins, at(motheduc=18) predict(outcome(3))
```

```
Predictive margins                                Number of obs   =           675
Model VCE      : OIM
Expression     : Pr(schooltype==Gymnasium), predict(outcome(3))
at             : motheduc           =           18
```

	Margin	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
---+---						
_cons	.9023163	.0326609	27.63	0.000	.838302	.9663305

Example: Determinants of secondary school choice (X)

In contrast, the estimation of e.g. the probability of the choice of Gymnasium for the maximum value of motheduc = 18 years at the means of the other individual characteristics with STATA leads to the following results:

```
margins, at((means)_all motheduc=18) predict(outcome(3))
```

```
Adjusted predictions                               Number of obs   =           675
Model VCE      : OIM
Expression     : Pr(schooltype==3), predict(outcome(3))
at             : motheduc           =           18
                mothlnlf           =    .5525926 (mean)
                loghhincome        =   11.05839 (mean)
                loghhsize          =    1.412881 (mean)
                birthorder         =     1.76 (mean)
                year1995           =    .1377778 (mean)
                year1996           =     .12 (mean)
                year1997           =    .1111111 (mean)
                year1998           =    .0888889 (mean)
                year1999           =    .1007407 (mean)
                year2000           =    .1037037 (mean)
                year2001           =    .1185185 (mean)
                year2002           =    .117037 (mean)
```

		Delta-method				
	Margin	Std. Err.	z	P> z	[95% Conf. Interval]	
__cons	.918859	.0294339	31.22	0.000	.8611696	.9765485

Example: Determinants of secondary school choice (XI)

The analysis of discrete changes of the choice probabilities for Hauptschule, Realschule, and Gymnasium due to a discrete change of motheduc requires the estimation of differences between the probabilities (an alternative for a discrete explanatory variable such as mothinf is the ML estimation with STATA by prefixing “i.” as well as the use of the commands as before, like “margins, dydx(mothinf) predict(outcome(3))”). For example, the average probabilities across all 675 pupils for several values of motheduc can be estimated. The following table reports the results:

motheduc (in years)	Hauptschule	Realschule	Gymnasium
7	0.6788	0.2465	0.0747
9	0.4966	0.3171	0.1864
10	0.3953	0.3339	0.2708
11	0.2975	0.3322	0.3703
12	0.2112	0.3121	0.4766
13	0.1418	0.2777	0.5805
14	0.0906	0.2353	0.6741
15	0.0555	0.1914	0.7531
16	0.0329	0.1506	0.8165
18	0.0108	0.0869	0.9023

Example: Determinants of secondary school choice (XII)

Interpretation:

- The increase from the minimum value of seven years to the maximum value of 18 years of education of the mother decreases the estimated average choice probabilities for Hauptschule and Realschule by 66.80 and 15.96 percentage points (from 0.6788 to 0.0108 and from 0.2465 to 0.0869), but increases the estimated average choice probability for Gymnasium by 82.76 percentage points (from 0.0747 to 0.9023). In the case of Gymnasium this means an immense increase of more than 1100%.
- The estimated change of the average choice probabilities for an increase of the years of education of the mother from nine to ten (which can be interpreted as the effect of “mittlere Reife”, i.e. the Realschule degree for the mother) is -10.13 percentage points for the case of Hauptschule and 8.44 percentage points for the case of Gymnasium
- The values for an increase of motheduc from ten to 13 years (which can be interpreted as the effect of “Abitur”, i.e. the Gymnasium degree for the mother) are -0.2535 for Hauptschule and 0.3097 for Gymnasium
- The values for an increase of motheduc from 13 to 16 years (which can be interpreted as the effect of an university degree for the mother) are -0.1089 for Hauptschule and 0.2360 for Gymnasium

3.3 Conditional logit models

Choice probabilities in conditional logit models ($i = 1, \dots, n; j = 1, \dots, J$):

$$p_{ij}(z_i, \gamma) = \frac{e^{\gamma' z_{ij}}}{\sum_{m=1}^J e^{\gamma' z_{im}}}$$

The inclusion of the ML estimator $\hat{\gamma}$ into these choice probabilities leads to the corresponding estimator $\hat{p}_{ij}(z_i, \hat{\gamma})$ of the choice probabilities for all categories $j = 1, \dots, J$.

Differences to (pure) multinomial logit models:

- The ML estimator $\hat{\gamma}$ is no longer choice-specific so that no normalization is necessary
- The estimators of the choice probabilities for alternative j do not only depend on the attributes z_{ij} , but also on all other alternative specific attributes in $z_i = (z'_{i1}, \dots, z'_{iJ})'$
- Since the alternative specific attributes vary across the categories and the observations, the ML estimation of conditional logit models with econometric software packages such as STATA requires another specific data organization (“long format” versus “wide format” before)

Example: Data organization in the conditional logit model

In order to examine the effect of the daily travel price (in Euro) and daily travel time (in minutes) on the choice between the use of car alone, carpool, bus, and train for the journey to work, the following table shows an exemplary data organization for the first three persons:

Person i	Transport modes	Choice	Travel price	Travel time
1	Car alone	0	6	50
1	Carpool	0	3	50
1	Bus	0	7	60
1	Train	1	9	30
2	Car alone	1	12	70
2	Carpool	0	4	70
2	Bus	0	7	90
2	Train	0	6	80
3	Car alone	0	3	20
3	Carpool	1	1	20
3	Bus	0	4	30
3	Train	0	5	20
⋮	⋮	⋮	⋮	⋮

Estimator of the (partial) marginal probability effect of a (continuous) alternative specific attribute z_{ijh} of alternative j on the choice of the same alternative j in conditional logit models ($i = 1, \dots, n, j = 1, \dots, J$):

$$\frac{\partial \hat{p}_{ij}(z_i, \hat{\gamma})}{\partial z_{ijh}} = \hat{p}_{ij}(z_i, \hat{\gamma}) [1 - \hat{p}_{ij}(z_i, \hat{\gamma})] \hat{\gamma}_h$$

Estimator of the (partial) marginal probability effect of a (continuous) alternative specific attribute z_{imh} of alternative m on the choice of another alternative j in conditional logit models ($i = 1, \dots, n, j = 1, \dots, J$):

$$\frac{\partial \hat{p}_{ij}(z_i, \hat{\gamma})}{\partial z_{imh}} = -\hat{p}_{ij}(z_i, \hat{\gamma}) \hat{p}_{im}(z_i, \hat{\gamma}) \hat{\gamma}_h \quad \forall m \neq j$$

In contrast to (pure) multinomial logit models, the sign of parameter estimators gives information about the direction of estimated marginal probability effects:

- If $\hat{\gamma}_h$ (e.g. the estimated parameter for price) is positive (negative), an increase of an attribute z_{ijh} for category j (e.g. price for bus) leads to an increase (decrease) of $\hat{p}_{ij}(z_i, \hat{\gamma})$ for the same category j (e.g. the estimated choice probability for bus)
- If $\hat{\gamma}_h$ (e.g. the estimated parameter for price) is positive (negative), an increase of an attribute z_{imh} for category m (e.g. price for train) leads to a decrease (increase) of $\hat{p}_{ij}(z_i, \hat{\gamma})$ for another category j (e.g. the estimated choice probability for bus)

Example: Determinants of fishing mode choice (I)

By using a conditional logit model, the effect of the following two alternative specific attributes on the choice between the four fishing modes charter (i.e. fishing on a charter boat), pier (i.e. fishing at the pier), private (i.e. fishing on a private boat), and beach (i.e. fishing on the beach) is examined on the basis of data from 1182 persons:

- Price (i.e. price of fishing mode in US dollars)
- Catchrate (i.e. average number of favorite fishes caught per hour by fishing mode)

In addition to such attributes, conditional logit models should generally include alternative specific constants in order to capture initial preferences for the different alternatives. Similar to the case of the parameters of individual characteristics in (pure) multinomial logit models, only $J-1$ alternative specific constants can be included so that category J is again the base category. The ML estimations of the conditional logit models (using beach as base category, respectively) lead to the following results (in line with the table on page 26, “fishmode” is a possible name for the identification of the four alternatives, “choice” is a possible name for the dependent variable, and “id” is a possible name for the identification of the persons in the sample):

Example: Determinants of fishing mode choice (II)

asclogit choice price, case(id) alternatives(fishmode) noconstant
basealternative(beach)

Alternative-specific conditional logit	Number of obs	=	4728
Case variable: id	Number of cases	=	1182
Alternative variable: fishmode	Alts per case: min	=	4
	avg	=	4.0
	max	=	4
	Wald chi2(1)	=	281.76
Log likelihood = -1372.9332	Prob > chi2	=	0.0000

choice	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
fishmode						
price	-.0179501	.0010694	-16.79	0.000	-.0200461	-.0158542

Example: Determinants of fishing mode choice (III)

asclogit choice price, case(id) alternatives(fishmode) basealternative(beach)

```
Alternative-specific conditional logit      Number of obs      =      4728
Case variable: id                        Number of cases     =      1182
Alternative variable: fishmode            Alts per case: min =          4
                                           avg   =          4.0
                                           max   =          4
                                           Wald chi2(1)      =      215.97
Log likelihood = -1236.7578                Prob > chi2        =      0.0000
```

choice		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
fishmode						
price		-.0252549	.0017185	-14.70	0.000	-.0286231 - .0218867
beach		(base alternative)				
charter						
_cons		1.669039	.1237266	13.49	0.000	1.426539 1.911538
pier						
_cons		.2839438	.1143707	2.48	0.013	.0597812 .5081063
private						
_cons		.8501593	.1138257	7.47	0.000	.6270651 1.073253

Example: Determinants of fishing mode choice (IV)

asclogit choice price catchrate, case(id) alternatives(fishmode) basealternative(beach)

```

Alternative-specific conditional logit      Number of obs      =      4728
Case variable: id                        Number of cases     =      1182
Alternative variable: fishmode            Alts per case: min =          4
                                           avg =          4.0
                                           max =          4
                                           Wald chi2(2)       =      229.35
Log likelihood = -1230.7838                Prob > chi2         =      0.0000
  
```

choice	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
fishmode						
price	-.0247896	.0017044	-14.54	0.000	-.0281301	-.021449
catchrate	.3771689	.1099707	3.43	0.001	.1616303	.5927074
-----+-----						
beach	(base alternative)					
-----+-----						
charter						
_cons	1.498888	.1329328	11.28	0.000	1.238345	1.759432
-----+-----						
pier						
_cons	.3070552	.1145738	2.68	0.007	.0824947	.5316158
-----+-----						
private						
_cons	.8713749	.1140428	7.64	0.000	.6478551	1.094895
-----+-----						

Example: Determinants of fishing mode choice (V)

An exemplary summary table for all estimation results has the following form

ML estimates (z statistics), dependent variable: fishing mode choice, base category: beach			
Explanatory variables	Model (1)	Model (2)	Model (3)
price	-0.018*** (-16.79)	-0.025*** (-14.70)	-0.025*** (-14.54)
catchrate	-	-	0.377*** (3.43)
constant: charter	-	1.669*** (13.49)	1.499*** (11.28)
constant: pier	-	0.284** (2.48)	0.307*** (2.68)
constant: private	-	0.850*** (7.47)	0.871*** (7.64)
Maximum value of log-likelihood function	-1372.93	-1236.76	-1230.78
Wald test statistic (all parameters)	281.76***	215.97***	229.35***

Note: *** (**, *) means that the appropriate parameter is different from zero or that the underlying null hypothesis is rejected at the 1% (5%, 10%) significance level, n = 1182

Example: Determinants of fishing mode choice (VI)

Interpretation:

- The price of a fishing mode j significantly decreases the probability of the choice of j (= estimated own price effect) and increases the probability of the choice of another fishing mode $m \neq j$ (= estimated cross price effect), *ceteris paribus*. Catchrate has a significantly positive effect on the own alternative.
- The initial preferences are significantly higher for charter, pier, and private relative to beach

Wald and likelihood ratio tests:

As an example, the null hypothesis that neither price nor catchrate has any effect on the fishing mode choice in model (3), i.e. that the two corresponding parameters are zero, is tested. The command for the Wald test in STATA is (this Wald test statistic is already reported in the underlying ML estimation with STATA since price and catchrate are the only explanatory variables so that the tested null hypotheses are identical):

```
test price=catchrate=0

( 1)  [fishmode]price - [fishmode]catchrate = 0
( 2)  [fishmode]price = 0

      chi2( 2) =    229.35
      Prob > chi2 =    0.0000
```

Example: Determinants of fishing mode choice (VII)

With respect to the application of the likelihood ratio test, the STATA command “estimates store unrestr” after the unrestricted ML estimation and the command “estimates store restr” after the restricted ML estimation are again necessary. The command for the likelihood ratio test in STATA is then:

```
lrtest unrestr restr
```

```
Likelihood-ratio test                    LR chi2(2)  =    533.88
(Assumption: restr nested in unrestr)    Prob > chi2 =    0.0000
```

Estimation of marginal probability effects:

- The estimation of average marginal probability effects is not directly possible with STATA (up to STATA 15)
- The STATA command “estat mfx” reports the estimated marginal probability effects at the means of the explanatory variables
- While this refers to all explanatory variables, the additional STATA command “varlist()” allows the limitation on a subset of explanatory variables
- The marginal probability effects can also be estimated at specific values of the explanatory variables

The estimation of marginal probability effects for price at the means of the explanatory variables in model (3) with STATA leads to the following results:

Example: Determinants of fishing mode choice (VIII)

```
estat mfx, varlist(price)
```

```
Pr(choice = beach|1 selected) = .0546071
```

variable	dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
price						
beach	-.00128	.00012	-10.66	0.000	-.001515 - .001044	103.42
charter	.000614	.00006	10.25	0.000	.000497 .000732	84.379
pier	.000098	.000017	5.88	0.000	.000065 .00013	103.42
private	.000568	.000056	10.16	0.000	.000458 .000677	55.257

```
Pr(choice = charter|1 selected) = .45376978
```

variable	dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
price						
beach	.000614	.00006	10.25	0.000	.000497 .000732	103.42
charter	-.006144	.000435	-14.12	0.000	-.006997 -.005291	84.379
pier	.000811	.000071	11.42	0.000	.000671 .00095	103.42
private	.00472	.000437	10.80	0.000	.003863 .005576	55.257

```
Pr(choice = pier|1 selected) = .07206028
```

variable	dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
price						
beach	.000098	.000017	5.88	0.000	.000065 .00013	103.42
charter	.000811	.000071	11.42	0.000	.000671 .00095	84.379
pier	-.001658	.000137	-12.07	0.000	-.001927 -.001389	103.42
private	.000749	.000066	11.30	0.000	.000619 .000879	55.257

Example: Determinants of fishing mode choice (IX)

Pr(choice = private|1 selected) = .41956284

variable	dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
price						
beach	.000568	.000056	10.16	0.000	.000458 .000677	103.42
charter	.00472	.000437	10.80	0.000	.003863 .005576	84.379
pier	.000749	.000066	11.30	0.000	.000619 .000879	103.42
private	-.006037	.000437	-13.82	0.000	-.006893 -.005181	55.257

Interpretation:

- At the means of the explanatory variables the estimated choice probabilities for the four fishing modes are $\hat{p}_{i1}(\bar{z}, \hat{\gamma}) = 0.0546$ for beach, $\hat{p}_{i2}(\bar{z}, \hat{\gamma}) = 0.4538$ for charter, $\hat{p}_{i3}(\bar{z}, \hat{\gamma}) = 0.0721$ for pier, and $\hat{p}_{i4}(\bar{z}, \hat{\gamma}) = 0.4196$ for private
- It follows e.g. for the estimated marginal probability effects of the price of private (charter) on the choice of private (at the means of the explanatory variables):

$$\hat{p}_{i4}(\bar{z}, \hat{\gamma})[1 - \hat{p}_{i4}(\bar{z}, \hat{\gamma})]\hat{\gamma}_1 = 0.4196 \cdot (1 - 0.4196) \cdot (-0.025) = -0.0060$$

$$-\hat{p}_{i4}(\bar{z}, \hat{\gamma})\hat{p}_{i2}(\bar{z}, \hat{\gamma})\hat{\gamma}_1 = -0.4196 \cdot 0.4538 \cdot (-0.025) = 0.0047$$

These values imply that an increase of the price of private (charter) by 1 dollar leads to an approximately estimated decrease (increase) of the choice probability for private by 0.60 (0.47) percentage points.

→ As already discussed above, general multinomial logit models can include both individual characteristics and alternative specific attributes as explanatory variables. In this case all previous interpretations from the (pure) multinomial and conditional logit models hold true. Similar to conditional logit models it is important to consider the specific data organization (“long format”).

Example: Data organization in the general multinomial logit model

The previous example of the analysis of the choice between the use of car alone, carpool, bus, and train for the journey to work now additionally includes the individual characteristic age (in years) as explanatory variable. The following table shows an exemplary data organization for the first two persons:

Person i	Transport modes	Choice	Travel price	Travel time	Age
1	Car alone	0	6	50	32
1	Carpool	0	3	50	32
1	Bus	0	7	60	32
1	Train	1	9	30	32
2	Car alone	1	12	70	51
2	Carpool	0	4	70	51
2	Bus	0	7	90	51
2	Train	0	6	80	51
⋮	⋮	⋮	⋮	⋮	⋮

Example: Determinants of fishing mode choice (I)

As in the previous example, the effect of price and catchrate on the choice between the four fishing modes charter, pier, private, and beach (base category) is examined on the basis of data from 1182 persons. However, the individual characteristic (monthly) income (in 1000 US dollars) is now (besides alternative specific constants) included as an additional explanatory variable. In such general multinomial logit models, all STATA commands as in the case of conditional logit models can be used. On the basis of the ML estimation of this specific multinomial logit model, the following tests and estimations are considered:

- The Wald test for the null hypothesis that the parameters for price and catchrate are zero
- The Wald test for the null hypothesis that the parameters for price and income are zero
- The corresponding likelihood ratio test for the null hypothesis that the parameters for price and income are zero (based on unrestricted and restricted ML estimations)
- The estimation of marginal probability effects for price and income at the means of the explanatory variables

The corresponding STATA commands lead to the following results:

Example: Determinants of fishing mode choice (II)

```
asclogit choice price catchrate, case(id) alternatives(fishmode) casevars(income)
basealternative(beach)
```

```
Alternative-specific conditional logit      Number of obs      =      4728
Case variable: id                          Number of cases    =      1182
Alternative variable: fishmode              Alts per case: min =         4
                                              avg =         4.0
                                              max =         4
                                              Wald chi2(5)      =      252.98
Log likelihood = -1215.1376                 Prob > chi2        =      0.0000
```

choice	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
fishmode						
price	-.0251166	.0017317	-14.50	0.000	-.0285106	-.0217225
catchrate	.357782	.1097733	3.26	0.001	.1426302	.5729337
-----+-----						
beach	(base alternative)					
-----+-----						
charter						
income	-.0332917	.0503409	-0.66	0.508	-.131958	.0653745
_cons	1.694366	.2240506	7.56	0.000	1.255235	2.133497
-----+-----						
pier						
income	-.1275771	.0506395	-2.52	0.012	-.2268288	-.0283255
_cons	.7779593	.2204939	3.53	0.000	.3457992	1.210119
-----+-----						
private						
income	.0894398	.0500671	1.79	0.074	-.0086898	.1875694
_cons	.5272788	.2227927	2.37	0.018	.0906132	.9639444
-----+-----						

Example: Determinants of fishing mode choice (III)

```
test price catchrate
```

```
( 1)  [fishmode]price = 0
( 2)  [fishmode]catchrate = 0
      chi2( 2) = 226.79
      Prob > chi2 = 0.0000
```

```
test price income
```

```
( 1)  [fishmode]price = 0
( 2)  [charter]income = 0
( 3)  [pier]income = 0
( 4)  [private]income = 0
      chi2( 4) = 235.66
      Prob > chi2 = 0.0000
```

```
estimates store unrestricted
```

```
asclogit choice catchrate, case(id) alternatives(fishmode) basealternative(beach)
```

```
estimates store restricted
```

```
lrtest unrestricted restricted
```

```
Likelihood-ratio test                    LR chi2(4) = 547.94
(Assumption: restricted nested in unrestricted) Prob > chi2 = 0.0000
```

Example: Determinants of fishing mode choice (IV)

```
estat mfx, varlist(price income)
```

```
Pr(choice = beach|1 selected) = .05248806
```

variable		dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
-----+-----							
price							
	beach	-.001249	.000121	-10.29	0.000	-.001487 - .001011	103.42
	charter	.000609	.000061	9.97	0.000	.000489 .000729	84.379
	pier	.000087	.000016	5.42	0.000	.000055 .000118	103.42
	private	.000553	.000056	9.88	0.000	.000443 .000663	55.257
-----+-----							
casevars							
	income	-.000721	.002319	-0.31	0.756	-.005266 .003823	4.0993

```
Pr(choice = charter|1 selected) = .46206853
```

variable		dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
-----+-----							
price							
	beach	.000609	.000061	9.97	0.000	.000489 .000729	103.42
	charter	-.006243	.000441	-14.15	0.000	-.007108 -.005378	84.379
	pier	.000764	.000071	10.69	0.000	.000624 .000904	103.42
	private	.00487	.000452	10.77	0.000	.003983 .005756	55.257
-----+-----							
casevars							
	income	-.021734	.00666	-3.26	0.001	-.034787 -.00868	4.0993

Example: Determinants of fishing mode choice (V)

Pr(choice = pier|1 selected) = .06584968

variable		dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
price							
	beach	.000087	.000016	5.42	0.000	.000055 .000118	103.42
	charter	.000764	.000071	10.69	0.000	.000624 .000904	84.379
	pier	-.001545	.000138	-11.16	0.000	-.001816 -.001274	103.42
	private	.000694	.000066	10.58	0.000	.000565 .000822	55.257
casevars							
	income	-.009306	.002719	-3.42	0.001	-.014635 -.003977	4.0993

Pr(choice = private|1 selected) = .41959373

variable		dp/dx	Std. Err.	z	P> z	[95% C.I.]	X
price							
	beach	.000553	.000056	9.88	0.000	.000443 .000663	103.42
	charter	.00487	.000452	10.77	0.000	.003983 .005756	84.379
	pier	.000694	.000066	10.58	0.000	.000565 .000822	103.42
	private	-.006117	.000444	-13.77	0.000	-.006987 -.005246	55.257
casevars							
	income	.031761	.006554	4.85	0.000	.018915 .044608	4.0993

3.4 More flexible multinomial discrete choice models

→ General multinomial logit models are the most widely used multinomial discrete choice models in empirical applications since the choice probabilities can be easily calculated due to their closed form. This allows the straightforward ML estimation and statistical testing in multinomial logit models.

Independence of Irrelevant Alternatives (IIA) in multinomial logit models:

This property implies that the choice probabilities between two alternatives (i.e. the odds) are independent of the existence of further alternatives. It has been developed in conditional logit models for the choice between the transport modes car, red bus, and blue bus and is based on the restrictive independence assumption of the error terms ε_{ij} . If the IIA property is not true, the multinomial logit model is misspecified so that the favorable properties of the ML estimator (consistency, asymptotic normality, asymptotic efficiency) become lost.

Hausman-McFadden test:

The idea of this test is that in the case of IIA the parameter estimates should not systematically change if some alternatives are omitted so that the same parameter estimates with and without some alternatives are compared. The test statistic (which is asymptotically χ^2 distributed with the number q of parameters as degrees of freedom under the null hypothesis of IIA) includes the difference of these estimates and the corresponding variance covariance matrixes. High values of the test statistic lead to the rejection of the null hypothesis.

Alternative multinomial discrete choice models:

- **Nested logit models:**
In these models the restrictive independence assumption across the extreme value distributed error terms ε_{ij} is weakened by grouping similar alternatives into nests (e.g. bus for red bus and blue bus). However, this model approach depends on the correct choice of the nests. Within the nests the IIA assumption still holds.
- **Mixed logit models:**
In these models the error terms ε_{ij} comprise two independent parts. The first part is independently and identically standard extreme value distributed as in the multinomial logit model. The second flexible part of error terms is able to allow any correlations and also heteroskedasticity. Due to this flexibility, the restrictive IIA property can be avoided. In contrast to multinomial and nested logit models, however, the choice probabilities are generally characterized by multiple integrals (see later) for which the calculation (as basis for the ML estimation) can be very difficult or even impossible with conventional deterministic numerical integration methods.
- **Multinomial probit models:**
In these models the error terms ε_{ij} are jointly normally distributed with an expectation vector zero and a flexible variance covariance matrix Σ . Different versions of multinomial probit models refer to different restrictions of Σ . Variance covariance parameters that are not normalized can be estimated.

Choice probabilities in multinomial probit models:

$$\begin{aligned}
 p_{ij}(X_i, \beta, \gamma) &= P\left(\varepsilon_{i1} - \varepsilon_{ij} < (\beta'_j X_i + \gamma' z_{ij}) - (\beta'_1 X_i + \gamma' z_{i1}); \dots; \right. \\
 &\quad \varepsilon_{i,j-1} - \varepsilon_{ij} < (\beta'_j X_i + \gamma' z_{ij}) - (\beta'_{j-1} X_i + \gamma' z_{i,j-1}); \\
 &\quad \varepsilon_{i,j+1} - \varepsilon_{ij} < (\beta'_j X_i + \gamma' z_{ij}) - (\beta'_{j+1} X_i + \gamma' z_{i,j+1}); \dots; \\
 &\quad \left. \varepsilon_{iJ} - \varepsilon_{ij} < (\beta'_j X_i + \gamma' z_{ij}) - (\beta'_J X_i + \gamma' z_{iJ}) \right) \\
 &= \int_{-\infty}^{(\beta'_j X_i + \gamma' z_{ij}) - (\beta'_1 X_i + \gamma' z_{i1})} \dots \int_{-\infty}^{(\beta'_j X_i + \gamma' z_{ij}) - (\beta'_{j-1} X_i + \gamma' z_{i,j-1})} \int_{-\infty}^{(\beta'_j X_i + \gamma' z_{ij}) - (\beta'_{j+1} X_i + \gamma' z_{i,j+1})} \dots \int_{-\infty}^{(\beta'_j X_i + \gamma' z_{ij}) - (\beta'_J X_i + \gamma' z_{iJ})} \\
 &\quad f_j(\omega_1, \dots, \omega_{j-1}, \omega_{j+1}, \dots, \omega_J) d\omega_1 \dots d\omega_{j-1} d\omega_{j+1} \dots d\omega_J
 \end{aligned}$$

$f_j(\cdot)$ is the joint density function of the normally distributed differences of the error terms ε_{ij} . In contrast to the simple form in (general) multinomial logit models, the choice probabilities have $(J-1)$ -dimensional integrals. This is the main practical problem of multinomial probit models with flexible correlation structures in the variance covariance matrix Σ . For a high number J of alternatives the computation of these integrals is not feasible with common integration methods in the iterative maximization process of the ML estimation. However, simulators can be used to approximate the probabilities which can be combined with the ML. This leads to a simulated ML (SML) estimation.