2. Flexible multinomial discrete choice models

2.1 Background

→ 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.

However, multinomial logit models are restrictive and limited in three directions:

- Multinomial logit models cannot be used for panel data analyses if unobserved heterogeneity is considered since this requires correlated error terms for each individual over time. Unobserved heterogeneity is also relevant in stated choice analyses, where generally (as in the case of panel data) several observations are considered for the same individuals.
- Multinomial logit models can represent systematic taste variation (i.e. taste variation with respect to observed characteristics of an individual), but not random taste heterogeneity (i.e. differences in tastes that are not related to observed characteristics)
- Multinomial logit models imply proportional substitution across the alternatives due to the Independence of Irrelevant Alternatives (IIA) property

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

- This property implies that the ratio of the choice probabilities between two alternatives j and r are independent of the existence of further alternatives:

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

- The property was developed in conditional logit models and shown 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 for the validity of IIA in multinomial logit models:

The idea of this test is that in the case of IIA, the parameter estimates should not systematically change if some alternatives are omitted. Thus, the same parameter estimates with and without some alternatives are compared.

Hausman-McFadden tests are specific versions of Hausman tests:

- Hausman tests are generally based on the difference between an estimated parameter vector $\hat{\beta}_a$ and another estimated parameter vector $\hat{\beta}_b$. Both estimators are consistent under the null hypothesis, but only $\hat{\beta}_b$ is efficient. In contrast, only $\hat{\beta}_a$, but not $\hat{\beta}_b$, is consistent under the alternative hypothesis.
- This procedure can be applied for the comparison of the estimated parameter vector $\hat{\beta}_{allalt}$ in a multinomial logit model with all J alternatives and $\hat{\beta}_{restralt}$ in a multinomial logit model with a restricted number of alternatives. Under H_0: IIA, $\hat{\beta}_{allalt}$ and $\hat{\beta}_{restralt}$ are consistent, but $\hat{\beta}_{allalt}$ is efficient. Under the alternative hypothesis (no IIA), $\hat{\beta}_{restralt}$ is consistent, but $\hat{\beta}_{allalt}$ is inconsistent.
- Therefore, the two estimators should not differ systematically under the null hypothesis. In contrast, if a statistical difference between the two estimators is found, it can be expected that IIA is not valid.
- Besides the difference between $\hat{\beta}_{allalt}$ and $\hat{\beta}_{restralt}$, also the estimated variance covariance matrix of this difference is included in the Hausman-McFadden test statistic:

$$HM = \left[\hat{\beta}_{restralt} - \hat{\beta}_{allalt}\right]' \left[\hat{Cov}\left(\hat{\beta}_{restralt} - \hat{\beta}_{allalt}\right)\right]^{-1} \left[\hat{\beta}_{restralt} - \hat{\beta}_{allalt}\right]$$

- Under H_0: IIA, HM is asymptotically χ^2 distributed with q (number of parameters) degrees of freedom. The null hypothesis is thus (for large n) rejected in favor of the alternative hypothesis at the significance level α if $HM > \chi^2_{q;1-\alpha}$.

Alternative multinomial discrete choice models:

- Nested logit models as example of generalized extreme value models:
  In these models the IIA assumption is weakened by grouping similar alternatives into nests (e.g. bus for red bus and blue bus). However, these models are still restrictive (e.g. within the nests the IIA assumption still holds), especially with respect to the other two limitations of multinomial logit models.
- Multinomial probit models:
  These models can avoid the three limitations of multinomial logit models by assuming that the error terms ε_{ij} are jointly normally distributed. In contrast to multinomial and nested logit models, however, the choice probabilities are characterized by multiple integrals, which must be completely simulated.
- Mixed logit models:
  These models can also avoid the three limitations of multinomial logit models by assuming that the random terms comprise two independent parts, i.e. one part that is independently and identically standard extreme value distributed and another part that is able to take up the three limitations. The choice probabilities also have to be simulated, but only partially due to the closed form of the first part of the error term.
- Latent class logit models:
  These models are related to mixed logit models, but use discrete distributions for the second part of the random terms instead of continuous distributions as in mixed logit models

2.2 Multinomial probit models

General multinomial probit models assume that the error terms ε_{ij} of the utility function u_{ij} are jointly normally distributed with expectation zero and variance covariance matrix Σ, i.e.:

$$\varepsilon_i = (\varepsilon_{i1},\ldots,\varepsilon_{iJ})' \sim N_J(0;\ \Sigma)$$

This leads to the following choice probabilities (i.e. probabilities that individual i chooses alternative j), where $\varphi_j(\cdot)$ is the joint density function of the normally distributed differences of the error terms ε_{ij}:

$$p_{ij}(X_i, \beta, \gamma, \Sigma) = P(y_{ij}=1|X_i, \beta, \gamma, \Sigma) = P(u_{ij} > u_{ij'};\ \forall j \neq j'|X_i, \beta, \gamma, \Sigma)$$

$$= \int_{-\infty}^{(\beta_j' x_i + \gamma' z_{ij})-(\beta_1' x_i + \gamma' z_{i1})} \cdots \int_{-\infty}^{(\beta_j' x_i + \gamma' z_{ij})-(\beta_{j-1}' x_i + \gamma' z_{i,j-1})} \int_{-\infty}^{(\beta_j' x_i + \gamma' z_{ij})-(\beta_{j+1}' x_i + \gamma' z_{i,j+1})} \cdots \int_{-\infty}^{(\beta_j' x_i + \gamma' z_{ij})-(\beta_J' x_i + \gamma' z_{iJ})}$$

$$\varphi_j(\omega_1,\ldots,\omega_{j-1},\omega_{j+1},\ldots,\omega_J)d\omega_1 \cdots d\omega_{j-1}d\omega_{j+1} \cdots d\omega_J$$

The probabilities can also be expressed with an indicator function $I(\cdot)$ that takes the value one when the statement in the parenthesis is true (i.e. that alternative j is chosen) and zero when it is false (i.e. another alternative is chosen):

$$p_{ij}(X_i, \beta, \gamma, \Sigma) =$$

$$\int I(y_{ij}=1|X_i, \beta, \gamma, \Sigma)\varphi_j(\omega_1,\ldots,\omega_{j-1},\omega_{j+1},\ldots,\omega_J)d\omega_1 \cdots d\omega_{j-1}d\omega_{j+1} \cdots d\omega_J$$

Some characteristics of multinomial probit models:

- The choice probabilities do not only depend on β and γ, but also on the components of the variance covariance matrix $\sum = (\sigma_{jj'})$, which contains 1/2·J(J+1) different variance and covariance parameters
- However, only 1/2·J·(J-1)-1 parameters, i.e. J-2 variance parameters and 1/2·(J-1)·(J-2) covariance parameters, can at most be formally identified
- In the most general case all these variance covariance parameters are estimated by mostly restricting the variances σ_{JJ} and $\sigma_{J-1,J-1}$ to the value one and the covariances σ_{jJ} ($\forall j \neq J$) to the value zero (so that e.g. two variances and three covariances can be estimated in the four-alternative probit model)
- In practice, the identification of the estimated variance covariance parameters can be very difficult or even impossible if the maximum number of parameters is included. Thus, more restrictions are often included such as the assumption that all variances are one or that all covariances are zero
- If both all variances are restricted to one and all covariances are restricted to zero, this leads to the independent multinomial probit model, which is very similar to and thus as restrictive as multinomial logit models
- The main characteristic is that the choice probabilities in flexible specifications generally have (J-1)-dimensional integrals. For large J, the computation of these integrals is not feasible with common numerical integration methods in the iterative maximization process of the ML estimation.

Choice probabilities with multiple integrals can be approximated by (stochastic) simulation methods, i.e. transformed draws of pseudorandom numbers:

- The main idea of simulation methods for approximating integrals is that integration over a density function is a form of averaging
  - First, several numbers of a random variable or vector are drawn from its underlying distribution
  - Second, a statistic within the integral that is based on this random variable or vector is calculated with the random draws, respectively
  - Third, the values are averaged across all random draws
- The basis for the calculation of choice probabilities in multinomial probit models is the J-1-dimensional normally distributed vector $\tilde{\varepsilon}_i = (\varepsilon_{i1} - \varepsilon_{ij}, \ldots, \varepsilon_{iJ} - \varepsilon_{ij})'$ of differences of the error terms (see slide 4, chapter 1)
- One obvious approach of simulating these choice probabilities refers to the second expression which is based on the indicator functions
  - First, several numbers $\tilde{\varepsilon}_i^r$ (r = 1,…, R) of the random vector $\tilde{\varepsilon}_i$ are drawn from a J-1-dimensional normal distribution by especially considering the modified variance covariance matrix of $\tilde{\varepsilon}_I$
  - Second, for each random draw it is determined whether the indictor function $I^r(\cdot)$ takes the value one (i.e. that the value of $\tilde{\varepsilon}_i^r$ falls into the interval A that is characterized by the integral bounds in the first expression) or alternatively the value zero

- Finally, it follows for this simple accept-reject or “crude frequency” simulator:

$$\tilde{p}_{ij}^{AR}(X_i, \beta, \gamma, \Sigma) = \frac{1}{R}\sum_{r=1}^{R} I^r(y_{ij}=1|X_i, \beta, \gamma, \Sigma)$$

- However, this simulation method has several shortcomings
  - It is possible that the simulator takes the value zero, which is highly problematic for the combination with estimation methods
  - The simulator is inefficient, i.e. a high number R of random draws is necessary for good approximations
  - The simulator is not smooth, i.e. changes in the parameters lead to discrete changes of the relative frequencies and thus approximated probabilities
- Many Monte Carlo experiments show that this simulator is worse than other simulators with respect to the approximation to correct probabilities
- For these reasons smoothed accept-reject simulators have been developed which replace the indicator function by a smooth and strictly positive function. Based on the same first step as the simple accept-reject simulator, the random numbers $\tilde{\varepsilon}_i^r$ (r = 1,…, R) are then included in this new function. Finally, the values are averaged across all R random numbers.
- However, several Monte Carlo experiments reveal the general superiority of the specific Geweke-Hajivassiliou-Keane (GHK) simulator that is also included in many econometric software packages such as Stata

GHK simulator:

- This simulator is the most important approach of “importance sampling” simulators, which are generally not based on draws from the normally distributed vector $\tilde{\varepsilon}_i$, but from another J-1-dimensional random vector V_i. This procedure should allow that all random draws fall into the interval A so that no draws are wasted (as in the case of the simple accept-reject simulator).
- The simulator is based on a representation of the choice probabilities p_{ij} as product of conditional probabilities. The random draws V_{ihr} (h = 1,…, J-1-1; r = 1,…, R) from the distribution of V_i are made sequentially for all components of V_i that follow a truncated standard normal distribution.
- For the simulated choice probabilities in multinomial probit models it follows with l_{km} (k, m = 1,…, J-1) as elements of the lower triangular matrix from the Cholesky decomposition of the variance covariance matrix of $\tilde{\varepsilon}_i$:

$$\tilde{p}_{ij}^{GHK}(X_i, \beta, \gamma, \Sigma) = \frac{1}{R}\sum_{r=1}^{R}\Phi\left(\frac{(\beta_j' x_i + \gamma' z_{ij}) - (\beta_1' x_i + \gamma' z_{i1})}{l_{11}}\right)\cdot$$

$$\Phi\left(\frac{(\beta_j' x_i + \gamma' z_{ij}) - (\beta_2' x_i + \gamma' z_{i2}) - l_{21} V_{i1r}}{l_{22}}\right)\cdots$$

$$\Phi\left(\frac{(\beta_j' x_i + \gamma' z_{ij}) - (\beta_J' x_i + \gamma' z_{iJ}) - l_{J1} V_{i1r} - l_{J2} V_{i2r} - \cdots - l_{J-1,J-2} V_{i,J-2,r}}{l_{J-1,J-1}}\right)$$

Remarks to the simulators:

- Just like the simple and smoothed accept-reject simulators, the GHK simulator is unbiased for the (choice) probabilities and its variance decreases as R increases
- In contrast to simple accept-reject and in line with smoothed accept-reject simulators, the GHK simulator is furthermore continuous and differentiable in the parameters of the choice probabilities and cannot take the value zero
- Monte Carlo experiments show that very small probabilities can be precisely approximated and that relatively small R are sufficient for an accurate approximation
- For the accuracy of the approximation, the quality of the drawn random numbers play an important role. The general basis are pseudorandom numbers which are directly drawn from the normal distribution or from the uniform distribution and then transformed to normally distributed pseudorandom numbers.
- However, the accuracy can be increased by using quasirandom sequences and thus including quasirandom numbers in the simulators. Prominent examples of such quasirandom sequences are Hammersley and especially Halton sequences, which are both (besides the common pseudorandom procedure) included in many econometric software packages such as Stata.

Simulated maximum likelihood method (SML):

- Due to the general problems or even impossibility of calculating multiple integrals, which are components of the log-likelihood function, the common ML (see Chapter 1) can often not be applied for estimating the parameters in these cases. However, the SML can be applied, which generally incorporates a simulator (such as the GHK simulator) into the ML approach.
- By including (e.g. GHK) simulated choice probabilities $\tilde{p}_{ij}(X_i, \beta, \gamma, \Sigma)$ in the case of multinomial probit models, it follows the specific simulated log-likelihood function and the SML estimator:

$$\log\tilde{L}(\beta, \gamma, \Sigma) = \sum_{i=1}^{n}\sum_{j=1}^{J} y_{ij}\log\tilde{p}_{ij}(X_i, \beta, \gamma, \Sigma)$$

$$\hat{\theta}_{SML} = \arg\max_{\theta}\left[\sum_{i=1}^{n}\sum_{j=1}^{J} y_{ij}\log\tilde{p}_{ij}(X_i, \beta, \gamma, \Sigma)\right]$$

- Properties of general SML estimators $\hat{\theta}_{SML}$:
  - For a fixed number of R, $\hat{\theta}_{SML}$ is inconsistent for $n \to \infty$
  - For R, $n \to \infty$, $\hat{\theta}_{SML}$ is consistent
  - In spite of this consistency, the asymptotic distribution of $\sqrt{n}(\hat{\theta}_{SML}-\theta)$ is biased if R does not increase sufficiently fast compared to n. As a consequence, $\hat{\theta}_{SML}$ is only asymptotically efficient if $\lim_{n \to \infty} \sqrt{n}/R = 0$.

Simulated classical tests:

- Due to the general problems or even impossibility of calculating multiple integrals, classical tests like Wald or likelihood ratio tests (see Chapter 1) can often not be applied in these cases. However, the inclusion of a simulator (such as the GHK simulator) and thus SML estimators $\hat{\theta}_{SML}$ into classical tests leads to simulated classical tests.
- It follows for the test statistics of the simulated Wald and likelihood ratio tests:

$$SWT = nc(\hat{\theta}_{SML})'\left[\frac{\partial c(\hat{\theta}_{SML})}{\partial \theta'}\hat{I}(\hat{\theta}_{SML})^{-1}\frac{\partial c(\hat{\theta}_{SML})'}{\partial \theta}\right]^{-1} c(\hat{\theta}_{SML})$$

$$SLRT = 2\left[\log L(\hat{\theta}_{SML,u}) - \log L(\hat{\theta}_{SML,r})\right]$$

- Properties of simulated classical tests
  – The tests are still asymptotically equivalent
  – However, SWT and SLRT are asymptotically noncentrally χ^2 distributed with q degrees of freedom under H_0 if $\lim_{n \to \infty} \sqrt{n}/R = c$
  – Only if $\lim_{n \to \infty} \sqrt{n}/R = 0$, SWT and SLRT are asymptotically centrally χ^2 distributed with q degrees of freedom under H_0

2.3 Mixed logit models

→ Mixed logit models are a highly flexible model class that can approximate any multinomial discrete choice model ("random utility model"). They are generally defined on the basis of their specific choice probabilities, but can be derived under several behavioral specifications, which provide specific interpretations, while being formally equivalent.

One derivation is the error components specification, which includes stochastic components that can create correlations across the alternatives. It follows for the utility function of i for alternative j:

$$u_{ij} = \beta_j' x_i^1 + \gamma' z_{ij}^1 + \beta_{ij}' x_i^2 + \gamma_i' z_{ij}^2 + \varepsilon_{ij}$$

As before, x_i^1, x_i^2, z_{ij}^1, and z_{ij}^2 are observed vectors of explanatory variables and β_j and γ are vectors of fixed parameters. However, β_{ij} and γ_i are vectors of error terms with expectation zero. Furthermore, the ε_{ij} are independently and identically standard extreme value distributed. As a consequence, the stochastic component of the utility function is now:

$$\eta_{ij} = \beta_{ij}' x_i^2 + \gamma_i' z_{ij}^2 + \varepsilon_{ij}$$

This specification allows correlations across the alternatives if x_i^2 and z_{ij}^2 are accordingly specified. In contrast, if x_i^2 and z_{ij}^2 are zero vectors, this leads to the standard multinomial logit model.

However, the most widely used derivation of mixed logit models in empirical applications is the random parameter specification. The basis is the following utility function of i for alternative j:

$$u_{ij} = \beta_{ij}^{'} x_i + \gamma_i^{'} z_{ij} + \varepsilon_{ij}$$

As in the case of the standard multinomial logit model, the ε_{ij} are independently and identically standard extreme value distributed. However, the parameter vectors β_{ij} and γ_i (summarized in δ_i) are no longer fixed, but vary across the individuals in the population and thus are continuous random vectors with specific density functions f(δ). These densities are functions of several parameters like the mean and variances of the components of δ in the population. Conditional on δ_i, it follows for the choice probabilities:

$$p_{ij}(X_i|\delta_i) = P(y_{ij}=1|X_i, \delta_i) = \frac{e^{\beta_{ij}^{'} x_i + \gamma_i^{'} z_{ij}}}{\sum_{m=1}^{J} e^{\beta_{im}^{'} x_i + \gamma_i^{'} z_{im}}}$$

However, δ_i is unknown so that it cannot be conditioned on δ. The unconditional choice probability is therefore the (mostly multiple) integral of $p_{ij}(X_i|\delta_i)$:

$$p_{ij}(X_i) = \int \frac{e^{\beta_j^{'} x_i + \gamma' z_{ij}}}{\sum_{m=1}^{J} e^{\beta_m^{'} x_i + \gamma' z_{im}}} f(\delta) d\delta$$

Parameters in mixed logit models :

- The first set of parameters refers to δ with density function f(δ)
- The second set of parameters refers to the vector θ that describes the distribution of δ (e.g. b and W if δ is normally distributed) so that it is useful to denote the density function as f(δ|θ)
- The choice probabilities are thus functions of θ, but do not depend on the values of δ, which are (similar to the other random terms ε_{ij}) integrated out
- In empirical applications, the estimated parameters in θ are mostly more interesting, even when sometimes the values of δ can also be of interest

According to the formula, the choice probabilities are weighted averages of the standard multinomial logit model probabilities, which are evaluated at different values of δ with weights that are given by f(δ|θ). According to the statistics literature, the distribution that provides the weights is called mixing distribution. If φ(δ|b, W) is the density function of normally distributed vectors with expectation b and variance covariance matrix W, it specifically follows:

$$p_{ij}(X_i, b, W) = \int \frac{e^{\beta_j' x_i + \gamma' z_{ij}}}{\sum_{m=1}^{J} e^{\beta_m' x_i + \gamma' z_{im}}} \varphi(\delta|b, W) d\delta$$

Other mixing distributions in empirical applications are lognormal, uniform, or triangular distributions.

Also the choice probabilities in mixed logit models can be approximated by simulation methods:

- The basis for the simulation of the probabilities is the general density function f(δ|θ) or specifically φ(δ|b, W) in the case of normal mixing distributions, from which a (pseudo)random number δ_r (r = 1,…, R) is drawn
- On this basis, the multinomial logit model formula for the rth draw can be calculated as follows:

$$\frac{e^{\beta'_{jr}x_i + \gamma'_r z_{ij}}}{\sum_{m=1}^{J} e^{\beta'_{mr}x_i + \gamma'_r{}' z_{im}}}$$

- After repeatedly drawing $\delta_1,\ldots,\delta_R$, and by averaging the corresponding multinomial logit model formulas, it follows for this mixed logit model simulator (in the case of normal mixing distributions):

$$\tilde{p}_{ij}(X_i, b, W) = \frac{1}{R}\sum_{r=1}^{R}\frac{e^{\beta'_{jr}x_i + \gamma'_r z_{ij}}}{\sum_{m=1}^{J} e^{\beta'_{mr}x_i + \gamma'_r{}' z_{im}}}$$

- This simulator is also unbiased for the (choice) probabilities, its variance decreases as R increases, it cannot take the value zero, and it is continuous and twice differentiable in the parameters of the choice probabilities

Further remarks to the simulator :

- In contrast to the simulation of the choice probabilities in multinomial probit models, which must be completely simulated, this specific mixed logit model simulator is only a partial simulation due to the closed form of the first part of the error term
- The mixed logit model simulator is in principle an accept-reject simulator that is based on draws from the random terms of multinomial discrete choice models and the calculation of the utilities from these draws. However, it is not a simple “crude frequency” simulator, but a logit-smoothed simulator, i.e. the calculated utilities are included into the multinomial logit model formula.
- Besides the use of pseudorandom numbers, which are directly drawn from $f(\delta|\theta)$ or specifically $\varphi(\delta|b, W)$, it is again possible to include quasirandomly drawn numbers in the mixed logit model simulator, especially based on Halton sequences
- Since the mixed logit model simulator is strictly positive and smooth, it can be easily included in the SML. It follows for the SML estimator:

$$\hat{\theta}_{SML} = \arg\max_{\theta} \left[\sum_{i=1}^{n} \sum_{j=1}^{J} y_{ij} \log \tilde{p}_{ij}(X_i, b, W) \right]$$

- On this basis, simulated classical tests such as simulated Wald and simulated likelihood ratio tests can be applied

Willingness to pay:

- The estimation of the WTP on the basis of mixed logit models is problematic if the parameters of both the interesting alternative specific attribute and the price or cost variables are random since the WTP is then the ratio of two randomly distributed terms. This can lead to WTP distributions that are strongly skewed and that may not even have defined moments.
- A common approach is thus to specify the parameter of the price or cost variable as fixed so that the estimated WTP is the ratio between the estimated mean of the parameter of an alternative specific attribute and the estimated price or cost parameter. However, the problem of this procedure is that it is often unreasonable to assume that all individuals have the same preferences for price or costs so that this approach is obviously a trade-off between reality and modelling convenience
- Alternative approaches with random parameters of price or cost variables: One approach considers the estimated means and variances of the random parameters and repeatedly draws from the assumed distributions. This leads to a number of WTP estimates (= ratios between the drawn parameter values), which are averaged across all draws. An alternative is the use of the so-called WTP space (instead of the preference space) approach, which specifies the distributions for the WTP directly so that problematic distributions for WTP are avoided.

2.4 Latent class logit models

→ Latent class logit models are related to mixed logit models: While mixed logit models include parameter vectors δ_i which are continuously distributed with the density function f(δ), the mixing distribution in latent class logit models is discrete. The application of latent class logit models is useful if there are Q segments in the population, each of which has its own preferences with respect to the choice among the alternatives.

For simplicity we only consider latent class logit models with alternative specific attributes in the following and thus the utility function of i for alternative j:

$$u_{ij} = \gamma_q' z_{ij} + \varepsilon_{ij}$$

Latent class logit models assume that the random vector γ can take the values $\gamma_1, \ldots, \gamma_Q$ with specific probabilities, which means that the individuals are implicitly sorted into a set of q = 1,…, Q classes. Since the other part ε_{ij} of the random terms is still independently and identically standard extreme value distributed, it follows for the choice probabilities under the condition that i belongs to class q:

$$p_{ijq}\left(z_i | \gamma_q\right) = \frac{e^{\gamma_q' z_{ij}}}{\sum_{m=1}^{J} e^{\gamma_q' z_{im}}}$$

Now $\gamma_q = (\gamma_{q1}, \ldots, \gamma_{qk_2})'$ is a class-specific vector of parameters for class q.

However, the probabilities that γ takes the values $\gamma_1, \ldots, \gamma_Q$ and thus the probabilities that i belongs to q are unknown. These probabilities are determined by a class membership model, where the membership to q depends on a vector x_i of individual characteristics. The corresponding unknown parameter vectors are $\beta_1, \ldots, \beta_Q$, where β_Q is normalized to the zero vector. By including the multinomial logit model formula in the class membership model, the probability that i belongs to class q is:

$$H_{iq} = \frac{e^{\beta_q' x_i}}{\sum_{q'=1}^{Q} e^{\beta_{q'}' x_i}}$$

It follows for the unconditional choice probability:

$$p_{ij}(z_i) = \sum_{q=1}^{Q} H_{iq} \frac{e^{\gamma_q' z_{ij}}}{\sum_{m=1}^{J} e^{\gamma_q' z_{im}}} = \sum_{q=1}^{Q} \left[\frac{e^{\beta_q' x_i}}{\sum_{q'=1}^{Q} e^{\beta_{q'}' x_i}} \frac{e^{\gamma_q' z_{ij}}}{\sum_{m=1}^{J} e^{\gamma_q' z_{im}}} \right]$$

These probabilities can then be included in the log-likelihood function so that it follows for the ML estimator:

$$\hat{\theta}_{ML} = \arg\max_{\theta} \left[\sum_{i=1}^{n} \sum_{j=1}^{J} y_{ij} \log \left(\sum_{q=1}^{Q} \left[\frac{e^{\beta_q' x_i}}{\sum_{q'=1}^{Q} e^{\beta_{q'}' x_i}} \frac{e^{\gamma_q' z_{ij}}}{\sum_{m=1}^{J} e^{\gamma_q' z_{im}}} \right] \right) \right]$$

Remarks to the parameter estimation:

- Since the choice probabilities are not characterized by (multiple) integrals, the parameters can generally be estimated by ML
- However, the maximization of the log-likelihood function refers to the Q structural parameter vectors $\gamma_1,\ldots,\gamma_Q$ and the Q-1 latent class parameter vectors $\beta_1,\ldots,\beta_{Q-1}$ so that the optimization problem is numerically not trivial compared to other common maximum likelihood estimations
- As a consequence, empirical applications with latent class logit models often use expectation-maximization (EM) algorithms, which are procedures for maximizing a log-likelihood function when standard procedures are numerically difficult or infeasible
- These algorithms are especially also included in many econometric software packages such as Stata

Remarks to the choice of the number Q of classes:

- One possibility is a contextual economic motivation of the class membership
- Another possibility is based on statistical measures to identify the latent class logit model with the best statistical fit such as the Bayesian information criterion (BIC) or the consistent Akaike information criterion (CAIC)
- Low values of these criteria indicate a better statistical fit so that the corresponding model with the lowest value of the information criteria are preferred in this respect

2.5 Applications

Example 1: Determinants of secondary school choice (I)

The determinants of the choice of 675 pupils in Germany between the three secondary school types Hauptschule, Realschule, and Gymnasium are again considered as in Chapter 1. However, as in Example 3, only the following individual characteristics are included:

- Years of education of the mother (motheduc)
- Dummy variable for labor force participation of the mother (mothinlf) that takes the value one if the mother is employed
- Logarithm of household income (loghhincome)

In the first step, the ML estimation of the multinomial logit model with and without robustly estimated variances of the estimated parameters on the basis of the two data organizations is again considered. Then, Hausman-McFadden tests (the standard Stata command is only possible on the basis of ML estimations without robustly estimated variances, while an additional command is possible, see later) for examining the validity of IIA are conducted (the additional command "alleqs" specifies that all equations of the models are included and not only the first equation as default). The corresponding Stata commands lead to the following results (the use of the "asclogit" would also be possible):

Example 1: Determinants of secondary school choice (II)

```
mlogit schooltype motheduc mothinlf loghhincome, base(1) robust

Multinomial logistic regression                   Number of obs   =        675
                                                  Wald chi2(6)    =      93.91
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -638.70703                 Pseudo R2       =     0.1285

------------------------------------------------------------------------------
             |               Robust
  schooltype |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Hauptschule  |  (base outcome)
-------------+----------------------------------------------------------------
Realschule   |
    motheduc |   .3425135   .0849103     4.03   0.000      .1760923    .5089347
    mothinlf |  -.1808054   .2113319    -0.86   0.392     -.5950083    .2333975
 loghhincome |   .1212413    .185545     0.65   0.513     -.2424202    .4849027
       _cons |  -4.878992   2.090943    -2.33   0.020     -8.977164   -.7808198
-------------+----------------------------------------------------------------
Gymnasium    |
    motheduc |   .6814142   .0916188     7.44   0.000      .5018447    .8609838
    mothinlf |   -.094855    .218878    -0.43   0.665     -.5238481    .3341381
 loghhincome |   1.169913   .2689158     4.35   0.000      .6428477    1.696978
       _cons |  -20.19249   3.042712    -6.64   0.000      -26.1561   -14.22889
------------------------------------------------------------------------------
```

Example 1: Determinants of secondary school choice (III)

```
mlogit schooltype motheduc mothinlf loghhincome, base(1)

Multinomial logistic regression                   Number of obs   =        675
                                                  LR chi2(6)      =     188.27
                                                  Prob > chi2     =     0.0000
Log likelihood = -638.70703                       Pseudo R2       =     0.1285

------------------------------------------------------------------------------
  schooltype |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Hauptschule  |  (base outcome)
-------------+----------------------------------------------------------------
Realschule   |
    motheduc |   .3425135   .0781934     4.38   0.000     .1892573    .4957698
    mothinlf |  -.1808054   .2091082    -0.86   0.387    -.5906499     .229039
 loghhincome |   .1212413   .1902865     0.64   0.524    -.2517133    .4941959
       _cons |  -4.878992   2.140196    -2.28   0.023      -9.0737   -.6842845
-------------+----------------------------------------------------------------
Gymnasium    |
    motheduc |   .6814142   .0804555     8.47   0.000     .5237243    .8391042
    mothinlf |   -.094855   .2151664    -0.44   0.659    -.5165735    .3268634
 loghhincome |   1.169913   .2383755     4.91   0.000     .7027056     1.63712
       _cons |  -20.19249   2.712505    -7.44   0.000    -25.50891   -14.87608
------------------------------------------------------------------------------

estimates store allalt
```

Example 1: Determinants of secondary school choice (IV)

ML estimation of the multinomial logit model by excluding the third alternative, i.e. Gymnasium:

```
mlogit schooltype motheduc mothinlf loghhincome if schooltype != 3, base(1)

Multinomial logistic regression                   Number of obs   =        398
                                                  LR chi2(3)      =      21.63
                                                  Prob > chi2     =     0.0001
Log likelihood = -265.05877                       Pseudo R2       =     0.0392

------------------------------------------------------------------------------
  schooltype |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Hauptschule  |  (base outcome)
-------------+----------------------------------------------------------------
Realschule   |
    motheduc |   .3332693   .0803386     4.15   0.000     .1758086    .4907301
    mothinlf |  -.1614151   .2099554    -0.77   0.442    -.5729201    .2500899
 loghhincome |   .1329202   .1977949     0.67   0.502    -.2547507    .5205911
       _cons |  -4.917706   2.223133    -2.21   0.027    -9.274966   -.5604462
------------------------------------------------------------------------------
```

Example 1: Determinants of secondary school choice (V)

```
hausman allalt, alleqs

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     allalt         .          Difference          S.E.
-------------+----------------------------------------------------------------
    motheduc |    .3425135     .3332693        .0092442                .
    mothinlf |   -.1808054    -.1614151       -.0193903                .
 loghhincome |    .1212413     .1329202       -.0116789                .
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from mlogit
           B = inconsistent under Ha, efficient under Ho; obtained from mlogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        0.80
                Prob>chi2 =      0.8484
                (V_b-V_B is not positive definite)
```

Interpretation:

The null hypothesis H_0: IIA cannot be rejected at common significance levels. However, in principle, all possible restricted numbers of alternatives must be analyzed (by using the same base alternative, respectively). If J = 3, this leads to three tests for three pairs of alternatives. The additional inclusion of the "constant" command would specify that also the constants are considered in the Hausman-McFadden test.

Example 1: Determinants of secondary school choice (VI)

Excluding the second alternative (Realschule) with the same base alternative:

```
mlogit schooltype motheduc mothinlf loghhincome if schooltype != 2, base(1)

hausman allalt, alleqs

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     allalt         .          Difference          S.E.
-------------+----------------------------------------------------------------
    motheduc |    .6814142     .6474181        .0339962               .
    mothinlf |    -.094855    -.0666842       -.0281708               .
 loghhincome |    1.169913     1.072782        .0971313               .
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from mlogit
           B = inconsistent under Ha, efficient under Ho; obtained from mlogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =    -4.00    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test
```

→ The value of the test statistic is negative, which is not possible theoretically, but which is interpreted as evidence that the null hypothesis H_0: IIA cannot be rejected. Such a result is not unusual when the sample is relatively small.

Example 1: Determinants of secondary school choice (VII)

Excluding the first alternative (Hauptschule) with the second alternative (Realschule) as base:

```
mlogit schooltype motheduc mothinlf loghhincome, base(2)

estimates store allalt

mlogit schooltype motheduc mothinlf loghhincome if schooltype != 1, base(2)

hausman allalt, alleqs

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     allalt         .           Difference          S.E.
-------------+----------------------------------------------------------------
    motheduc |    .3389007     .3548258       -.0159251                 .
    mothinlf |    .0859504     .0304384         .055512                 .
 loghhincome |    1.048672     1.041221        .0074502                 .
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from mlogit
           B = inconsistent under Ha, efficient under Ho; obtained from mlogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =    -2.55    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test
```

→ In sum, there is no clear evidence for or against IIA

Example 1: Determinants of secondary school choice (VIII)

Now, the ML and SML estimations of the independent multinomial probit model with robustly estimated variances of the estimated parameters on the basis of the two data organizations are considered. The corresponding Stata commands lead to the following results:

```
mprobit schooltype motheduc mothinlf loghhincome, base(1) robust

Multinomial probit regression                     Number of obs   =        675
                                                  Wald chi2(6)    =      98.84
Log pseudolikelihood = -640.76755                 Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |               Robust
  schooltype |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Hauptschule  |  (base outcome)
-------------+----------------------------------------------------------------
Realschule   |
    motheduc |   .2280785   .0649979     3.51   0.000      .100685    .3554719
    mothinlf |  -.1002382   .1654484    -0.61   0.545    -.4245112    .2240348
 loghhincome |   .0899959   .1488687     0.60   0.545    -.2017813    .3817731
       _cons |  -3.384255   1.659129    -2.04   0.041    -6.636088   -.1324219
-------------+----------------------------------------------------------------
Gymnasium    |
    motheduc |   .4946286   .0702575     7.04   0.000     .3569265    .6323307
    mothinlf |  -.0515035   .1691124    -0.30   0.761    -.3829577    .2799506
 loghhincome |   .8545471   .1993188     4.29   0.000     .4638894    1.245205
       _cons |  -14.72041   2.235405    -6.59   0.000    -19.10172   -10.33909
------------------------------------------------------------------------------
```

Example 1: Determinants of secondary school choice (IX)

```
asmprobit choice, case(id) alternatives(alt) casevars(motheduc mothinlf loghhincome)
base(1) correlation(independent) stddev(homoskedastic) robust

Alternative-specific multinomial probit        Number of obs      =       2025
Case variable: id                              Number of cases    =        675

Alternative variable: alt                      Alts per case: min =          3
                                                              avg =        3.0
                                                              max =          3
Integration sequence:           Hammersley
Integration points:                    150        Wald chi2(6)    =      98.67
Log simulated-pseudolikelihood = -640.78024       Prob > chi2     =     0.0000
                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1            |  (base alternative)
-------------+----------------------------------------------------------------
2            |
    motheduc |    .228087   .0650248     3.51   0.000     .1006408    .3555332
    mothinlf |  -.1002515   .1654012    -0.61   0.544    -.4244319    .2239288
 loghhincome |   .0900133   .1488548     0.60   0.545    -.2017367    .3817634
       _cons |  -3.384533   1.658794    -2.04   0.041     -6.63571   -.1333564
-------------+----------------------------------------------------------------
3            |
    motheduc |   .4946415   .0703171     7.03   0.000     .3568227    .6324604
    mothinlf |  -.0515079   .1690588    -0.30   0.761     -.382857    .2798412
 loghhincome |   .8545628   .1992574     4.29   0.000     .4640255      1.2451
       _cons |  -14.72071   2.235194    -6.59   0.000    -19.10161   -10.33981
------------------------------------------------------------------------------
```

Example 1: Determinants of secondary school choice (X)

Interpretation:

- The estimation results are all qualitatively very similar and especially very similar to the estimation results in multinomial logit models. The differences in the parameter estimates in the multinomial logit and independent multinomial probit models are based on different scale factors and thus on the different distribution assumptions in the two approaches.
- While the “mprobit” command leads to ML estimations on the basis of a numerical integration method (which is possible in independent multinomial probit models since there are no multiple integrals), the “asmprobit” command leads to SML estimations. As a consequence, the estimation results are quantitatively very similar, but not identical.
- Due to the additional Stata commands “correlation(independent)” and “stddev(homoskedastic)”, the variances of the variance covariance matrix ∑ are restricted to the value one and the covariances are restricted to the value zero, which is in line with independent multinomial probit models
- Generally, it is also possible to only include one of the two commands or none of the command to specify ∑. In the latter case, this leads to the most flexible multinomial probit model.

Example 1: Determinants of secondary school choice (XI)

- Generally, the SML estimation of flexible multinomial probit models can be very unstable so that often the simulated log-likelihood function does not converge to a maximum due to flat regions of the function (see also later). Therefore, some restrictions with respect to $\sum$ can become necessary. In fact, the SML estimation of flexible multinomial probit models with only individual characteristics is even not possible with the "mprobit" command and commonly does not converge to a maximum in the optimization process on the basis of the "asmprobit" command.
- With respect to the random numbers in the SML estimation (only on the basis of the "asmprobit" command), the default is the use of quasirandom Hammersley sequences. Pseudorandom numbers and quasirandom numbers on the basis of Halton sequences can be included by the additional Stata commands "intmethod(random)" and "intmethod(halton)".
- With respect to the number R of random draws, the defaults are $R = 50 \cdot J$ for Hammersley and Halton sequences and $R = 100 \cdot J$ for pseudorandom numbers. The number R can be manually determined by including the additional command "intpoints(#)", e.g. "intpoints(1000) for $R = 1000$ random draws. In line with the asymptotic properties of the SML, higher R generally provide more reliable estimation results, but require additional computation time.

Example 1: Determinants of secondary school choice (XII)

Wald and likelihood ratio tests of the null hypothesis that motheduc has no effect on the secondary school choice on the basis of the ML estimations (the corresponding SML estimations with the "asmprobit" command would also be possible) without robustly estimated variances (likelihood ratio tests in the case of robustly estimated variances are only possible if the additional command "force" is included, see later):

```
mprobit schooltype motheduc mothinlf loghhincome, base(1)

estimates store unrestricted

test motheduc

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

           chi2(  2) =   88.03
         Prob > chi2 =    0.0000

mprobit schooltype mothinlf loghhincome, base(1)

estimates store restricted

lrtest unrestricted restricted

Likelihood-ratio test                                 LR chi2(2)  =    117.33
(Assumption: restricted nested in unrestricted)       Prob > chi2 =    0.0000
```

Example 1: Determinants of secondary school choice (XIII)

The estimation of the average marginal probability effect of motheduc on the choice of Gymnasium (an estimation at the means of the individual characteristics would also be possible) on the basis of the ML estimations (the corresponding SML estimations with the "asmprobit" command are not possible) without robustly estimated variances with Stata leads to the following results:

```
mprobit schooltype motheduc mothinlf loghhincome, base(1)

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

Average marginal effects                          Number of obs   =        675
Model VCE    : OIM
Expression   : Pr(schooltype==Gymnasium), predict(outcome(3))
dy/dx w.r.t. : motheduc
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    motheduc |   .0904379   .0081469    11.10   0.000     .0744702    .1064057
------------------------------------------------------------------------------
```

This value of 0.0904 (which is identical to the case of robustly estimated variances, i.e. only the z-value differs) 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 9.04 percentage points. This estimated effect is very similar to the value in the multinomial logit model.

Example 1: Determinants of secondary school choice (XIV)

The estimation of the average probabilities of the choice of Gymnasium for motheduc = 7 and motheduc = 18 years with Stata (as basis for the estimation of discrete probability effects) leads to the following results (as in the case of "asclogit", no "margins" command, but only the "mfx" command is possible after an "asmprobit" command, see later):

```
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

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .0650831   .0192294     3.38   0.001     .0273942    .1027721
------------------------------------------------------------------------------

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

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .9034587   .0344423    26.23   0.000     .8359531    .9709643
------------------------------------------------------------------------------
```

Example 1: Determinants of secondary school choice (XV)

Now, SML estimations of a mixed logit model are considered. The “mixlogit” command requires the same data organization as in the case of the “asclogit” and “asmprobit” commands. Individual characteristics cannot be simply included in the command, but have to be interacted with each alternative specific constant except for the base alternative. Therefore, the following explanatory variables are included:

- Interaction terms of years of education of the mother with Realschule (motheduc_real) and with Gymnasium (motheduc_gym)
- Interaction terms of labor force participation of the mother with Realschule (mothinlf_real) and with Gymnasium (mothinlf_gym)
- Interaction terms of the logarithm of household income with Realschule (loghhincome_real) and with Gymnasium (loghhincome_gym)
- Alternative specific constants for Realschule and Gymnasium

Furthermore, it must be specified, which parameters are assumed to be fixed or random. Usually, parameters of interaction terms are considered as fixed. SML estimations with and without robustly estimated variances of the estimated parameters lead to the following results (“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 1: Determinants of secondary school choice (XVI)

```
mixlogit choice motheduc_real mothinlf_real loghhincome_real motheduc_gym mothinlf_gym
loghhincome_gym, group(id) rand(realschule gymnasium) robust

Mixed logit model                                 Number of obs   =       2025
                                                  Wald chi2(8)    =      88.58
Log likelihood = -638.58654                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------------
                 |               Robust
          choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------+------------------------------------------------------------------
Mean             |
   motheduc_real |   .3424605   .0861826     3.97   0.000     .1735457    .5113754
   mothinlf_real |  -.1816103    .215202    -0.84   0.399    -.6033985    .2401779
loghhincome_real |   .1125921   .1917427     0.59   0.557    -.2632166    .4884008
    motheduc_gym |   .6938723   .0973739     7.13   0.000      .503023    .8847217
    mothinlf_gym |  -.0925259   .2235057    -0.41   0.679    -.5305891    .3455373
 loghhincome_gym |   1.197237   .2755734     4.34   0.000     .6571231    1.737351
      realschule |   -4.80257   2.150137    -2.23   0.026    -9.016761   -.5883794
       gymnasium |  -20.64829   3.172731    -6.51   0.000    -26.86673   -14.42985
-----------------+------------------------------------------------------------------
SD               |
      realschule |   .3386067   .5859615     0.58   0.563    -.8098567     1.48707
       gymnasium |   .3588516   .4394142     0.82   0.414    -.5023843    1.220088
------------------------------------------------------------------------------------
```

→ The reported simulated Wald test statistic (if the "robust" command is included) refers to the null hypothesis that the eight mean parameters are zero. The hypothesis can be rejected at all common significance levels.

Example 1: Determinants of secondary school choice (XVII)

```
mixlogit choice motheduc_real mothinlf_real loghhincome_real motheduc_gym mothinlf_gym
loghhincome_gym, group(id) rand(realschule gymnasium)

Mixed logit model                                 Number of obs   =       2025
                                                  LR chi2(2)      =       0.24
Log likelihood = -638.58654                       Prob > chi2     =     0.8865

------------------------------------------------------------------------------------
          choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------+------------------------------------------------------------------
Mean             |
   motheduc_real |   .3424605    .079361     4.32   0.000     .1869158    .4980053
   mothinlf_real |  -.1816103   .2130567    -0.85   0.394    -.5991938    .2359732
loghhincome_real |   .1125921   .2009602     0.56   0.575    -.2812826    .5064669
    motheduc_gym |   .6938723   .1027571     6.75   0.000     .4924722    .8952725
    mothinlf_gym |  -.0925259    .219724    -0.42   0.674     -.523177    .3381252
 loghhincome_gym |   1.197237    .279466     4.28   0.000     .6494937     1.74498
      realschule |   -4.80257   2.234783    -2.15   0.032    -9.182663   -.4224768
       gymnasium |  -20.64829   3.589617    -5.75   0.000    -27.68381   -13.61277
-----------------+------------------------------------------------------------------
SD               |
      realschule |   .3386067   1.088697     0.31   0.756      -1.7952    2.472413
       gymnasium |   .3588516   .9569633     0.37   0.708    -1.516762    2.234465
------------------------------------------------------------------------------------
```

→ The reported simulated likelihood ratio test statistic refers to the null hypothesis that the two parameters of the standard deviations are zero. The hypothesis cannot be rejected at common significance levels.

Example 1: Determinants of secondary school choice (XVIII)

Interpretation:

- The estimation results for the three individual characteristics are qualitatively very similar to the former results
- There is no indication of random alternative specific constants since their standard deviations are not different from zero at any common significance level and due to the reported simulated likelihood ratio test statistic
- With respect to the random draws for the simulation of the choice probabilities, quasirandomly drawn numbers based on Halton sequences are considered. While the default is R = 50, the number R of Halton draws can be individually determined by including the additional command "nrep(#)".
- With respect to the distribution of all random parameters, the default is that they are normally distributed and independent. By including the additional command "ln(#)", the random parameters for the last # variables are assumed to be log-normally distributed. By including "corr", correlations between the random parameters are modeled and the estimated elements of the lower triangular matrix of the Cholesky decomposition of the variance covariance matrix (W for the normal distribution) are reported.

Example 1: Determinants of secondary school choice (XIX)

The SML estimation of the mixed logit model with correlated, normally distributed alternative specific constants, R = 500 Halton draws, and robustly estimated variances of the estimated parameters leads to the following results:

```
mixlogit choice motheduc_real mothinlf_real loghhincome_real motheduc_gym mothinlf_gym
loghhincome_gym, group(id) rand(realschule gymnasium) corr nrep(500) robust

Mixed logit model                                  Number of obs    =       2025
                                                   Wald chi2(8)     =     107.26
Log likelihood = -638.70619                        Prob > chi2      =     0.0000

------------------------------------------------------------------------------------
                 |               Robust
          choice |      Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
-----------------+------------------------------------------------------------------
   motheduc_real |   .3424921   .0849146     4.03   0.000     .1760625    .5089217
   mothinlf_real |  -.1808106   .2113348    -0.86   0.392    -.5950192     .233398
loghhincome_real |   .1211804   .1855585     0.65   0.514    -.2425076    .4848684
    motheduc_gym |    .681526   .0916234     7.44   0.000     .5019474    .8611046
    mothinlf_gym |  -.0948407   .2189197    -0.43   0.665    -.5239155     .334234
 loghhincome_gym |   1.170197   .2689462     4.35   0.000     .6430718    1.697321
      realschule |    -4.8782   2.091143    -2.33   0.020    -8.976765   -.7796358
       gymnasium |   -20.1971    3.04281    -6.64   0.000     -26.1609    -14.2333
-----------------+------------------------------------------------------------------
            /l11 |   .0100089    .053667     0.19   0.852    -.0951765    .1151944
            /l21 |  -.0271508   .0543663    -0.50   0.617    -.1337067    .0794052
            /l22 |   .0213946   .0480394     0.45   0.656    -.0727608      .11555
------------------------------------------------------------------------------------
```

Example 1: Determinants of secondary school choice (XX)

If the “robust” command is not included, a simulated likelihood ratio test statistic instead of a simulated Wald test statistic is again reported, which now refers to the null hypothesis that the three variance covariance parameters (i.e. two variance parameters and one correlation parameter) are zero:

```
mixlogit choice motheduc_real mothinlf_real loghhincome_real motheduc_gym mothinlf_gym
loghhincome_gym, group(id) rand(realschule gymnasium) corr nrep(500)
```

```
Mixed logit model                                 Number of obs   =       2025
                                                  LR chi2(3)      =       0.00
Log likelihood = -638.70619                       Prob > chi2     =     1.0000

------------------------------------------------------------------------------
          choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
   motheduc_real |   .3424921   .0782362     4.38   0.000     .1891519    .4958323
   mothinlf_real |  -.1808106   .2091121    -0.86   0.387    -.5906628    .2290416
loghhincome_real |   .1211804   .1903935     0.64   0.524    -.2519839    .4943448
    motheduc_gym |    .681526   .0807363     8.44   0.000     .5232859    .8397662
    mothinlf_gym |  -.0948407   .2152063    -0.44   0.659    -.5166373    .3269559
 loghhincome_gym |   1.170197   .2388834     4.90   0.000     .7019938    1.638399
      realschule |    -4.8782   2.141877    -2.28   0.023    -9.076202   -.6801984
       gymnasium |   -20.1971   2.724176    -7.41   0.000    -25.53639   -14.85782
-----------------+------------------------------------------------------------
            /l11 |   .0100089    1.10783     0.01   0.993    -2.161298    2.181316
            /l21 |  -.0271508    1.14021    -0.02   0.981    -2.261921     2.20762
            /l22 |   .0213946   .9595362     0.02   0.982    -1.859262    1.902051
------------------------------------------------------------------------------
```

Example 1: Determinants of secondary school choice (XXI)

Simulated Wald test and simulated likelihood ratio test of the null hypothesis that motheduc has no effect on the secondary school choice on the basis of the first SML estimation:

```
mixlogit choice motheduc_real mothinlf_real loghhincome_real motheduc_gym mothinlf_gym
loghhincome_gym, group(id) rand(realschule gymnasium)

test motheduc_real motheduc_gym

 ( 1)  [Mean]motheduc_real = 0
 ( 2)  [Mean]motheduc_gym = 0

           chi2(  2) =   45.84
         Prob > chi2 =    0.0000

estimates store unrestricted

mixlogit choice mothinlf_real loghhincome_real mothinlf_gym loghhincome_gym, group(id)
rand(realschule gymnasium)

estimates store restricted

lrtest unrestricted restricted, force

Likelihood-ratio test                                 LR chi2(2)  =    117.79
(Assumption: restricted nested in unrestricted)       Prob > chi2 =    0.0000
```

→ The additional command “force” allows calculations of the likelihood ratio test statistic where “lrtest” would not allow it like here in mixed logit models, but also in other cases like the inclusion of robustly estimated variances

---

Example 2: Determinants of travel mode choice (I)

The determinants of the choice between air (base category), train, bus, and car are again examined as in Example 2 in Chapter 1 by including the following alternative specific attributes and individual characteristic:

- Travelcost (i.e. generalized costs of travel in US dollars, which is equal to the sum of in vehicle cost and the product of travel time and the value of travel time savings)
- Termtime (i.e. terminal time in minutes, which is zero for car transportation)
- Household income (in 1000 US dollars)

In the first step, the ML estimation of the multinomial logit model with and without robustly estimated variances of the estimated parameters is again considered. Then, a Hausman-McFadden test on the basis of ML estimations without robustly estimated variances (however, the additional command “force” would again allow the calculation of this test statistic if robustly estimated variances are included) for examining the validity of IIA are conducted. The corresponding Stata commands lead to the following results:

Example 2: Determinants of travel mode choice (II)

```
asclogit choice travelcost termtime, case(id) alternatives(travelmode) casevars(income) robust

Alternative-specific conditional logit         Number of obs      =        840
Case variable: id                              Number of cases    =        210

Alternative variable: travelmode               Alts per case: min =          4
                                                              avg =        4.0
                                                              max =          4

                                                  Wald chi2(5)    =      74.24
Log pseudolikelihood = -189.52515                 Prob > chi2     =     0.0000

                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
travelmode   |
  travelcost |  -.0109274   .0049767    -2.20   0.028    -.0206815   -.0011732
    termtime |  -.0954606    .014622    -6.53   0.000    -.1241191    -.066802
-------------+----------------------------------------------------------------
air          |  (base alternative)
-------------+----------------------------------------------------------------
train        |
      income |  -.0511884   .0153147    -3.34   0.001    -.0812046   -.0211722
       _cons |  -.3249561   .6562138    -0.50   0.620    -1.611111    .9611992
-------------+----------------------------------------------------------------
bus          |
      income |  -.0232107    .013753    -1.69   0.091    -.0501661    .0037447
       _cons |  -1.744529   .6312441    -2.76   0.006    -2.981745   -.5073138
-------------+----------------------------------------------------------------
car          |
      income |   .0053735   .0099531     0.54   0.589    -.0141343    .0248813
       _cons |  -5.874813   .9180023    -6.40   0.000    -7.674065   -4.075562
------------------------------------------------------------------------------
```

Example 2: Determinants of travel mode choice (III)

```
asclogit choice travelcost termtime, case(id) alternatives(travelmode) casevars(income)

Alternative-specific conditional logit         Number of obs      =        840
Case variable: id                              Number of cases    =        210

Alternative variable: travelmode               Alts per case: min =          4
                                                              avg =        4.0
                                                              max =          4

                                                  Wald chi2(5)    =     105.78
Log likelihood = -189.52515                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
travelmode   |
  travelcost |  -.0109274   .0045878    -2.38   0.017    -.0199192   -.0019355
    termtime |  -.0954606   .0104732    -9.11   0.000    -.1159876   -.0749335
-------------+----------------------------------------------------------------
air          |  (base alternative)
-------------+----------------------------------------------------------------
train        |
      income |  -.0511884   .0147352    -3.47   0.001    -.0800689   -.0223079
       _cons |  -.3249561   .5763335    -0.56   0.573    -1.454549    .8046369
-------------+----------------------------------------------------------------
bus          |
      income |  -.0232107   .0162306    -1.43   0.153     -.055022    .0086006
       _cons |  -1.744529   .6775004    -2.57   0.010    -3.072406   -.4166531
-------------+----------------------------------------------------------------
car          |
      income |   .0053735   .0115294     0.47   0.641    -.0172237    .0279707
       _cons |  -5.874813   .8020903    -7.32   0.000    -7.446882   -4.302745
------------------------------------------------------------------------------

estimates store allalt
```

Example 2: Determinants of travel mode choice (IV)

ML estimation by excluding the fourth alternative, i.e. car:

```
asclogit choice travelcost termtime if travelmode != 4, case(id) alternatives(travelmode)
casevars(income)

Alternative-specific conditional logit         Number of obs      =        453
Case variable: id                              Number of cases    =        151

Alternative variable: travelmode               Alts per case: min =          3
                                                              avg =        3.0
                                                              max =          3

                                                  Wald chi2(4)    =      65.37
Log likelihood = -96.171779                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
travelmode   |
  travelcost |  -.0118851   .0060469    -1.97   0.049    -.0237367   -.0000334
    termtime |  -.0797208   .0106985    -7.45   0.000    -.1006894   -.0587521
-------------+----------------------------------------------------------------
air          |  (base alternative)
-------------+----------------------------------------------------------------
train        |
      income |  -.0565631   .0150837    -3.75   0.000    -.0861267   -.0269996
       _cons |   .1598394   .5569027     0.29   0.774    -.9316697    1.251349
-------------+----------------------------------------------------------------
bus          |
      income |  -.0297777   .0161852    -1.84   0.066    -.0615002    .0019448
       _cons |  -1.283321   .6551554    -1.96   0.050    -2.567402      .00076
------------------------------------------------------------------------------
```

Example 2: Determinants of travel mode choice (V)

```
hausman allalt, alleqs

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     allalt         .          Difference          S.E.
-------------+----------------------------------------------------------------
travelmode   |
  travelcost |   -.0109274    -.0118851        .0009577               .
    termtime |   -.0954606    -.0797208       -.0157398               .
-------------+----------------------------------------------------------------
train        |
      income |   -.0511884    -.0565631        .0053748               .
-------------+----------------------------------------------------------------
bus          |
      income |   -.0232107    -.0297777         .006567        .0012122
------------------------------------------------------------------------------
                         b = consistent under Ho and Ha; obtained from asclogit
          B = inconsistent under Ha, efficient under Ho; obtained from asclogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(4) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       20.92
                Prob>chi2 =      0.0003
                (V_b-V_B is not positive definite)
```

→ Here, the null hypothesis H_0: IIA can be rejected at low significance levels so that more flexible multinomial discrete choice models such as multinomial probit, mixed logit, and/or latent class logit models should be used

---

Example 2: Determinants of travel mode choice (VI)

Now, the following estimations and tests in different multinomial probit models are considered:

- SML estimation of the independent multinomial probit model (with robustly estimated variances)
- SML estimation of a more flexible multinomial probit model (i.e. the variances of ∑ are restricted to the value one) (with robustly estimated variances) whereby the quasirandom numbers are based on Halton sequences with R = 1000
- SML estimation of the most flexible multinomial probit model (without robustly estimated variances)
- SML estimation of the most flexible multinomial probit model (without robustly estimated variances), whereby not the estimated components of the variance covariance matrix of $\tilde{\varepsilon}_i = (\varepsilon_{i1} - \varepsilon_{ij}, \ldots, \varepsilon_{iJ} - \varepsilon_{ij})'$ are reported (which is the default), but the estimated variances and covariances of ∑
- The estimation of marginal probability effects for travelcost and income at the means of the explanatory variables (with robustly estimated variances)

The corresponding Stata commands lead to the following results:

Example 2: Determinants of travel mode choice (VII)

```
asmprobit choice travelcost termtime, case(id) alternatives(travelmode) casevars(income)
correlation(independent) stddev(homoskedastic) robust

Alternative-specific multinomial probit        Number of obs      =        840
Case variable: id                              Number of cases    =        210

Alternative variable: travelmode               Alts per case: min =          4
                                                              avg =        4.0
                                                              max =          4

Integration sequence:           Hammersley
Integration points:                    200        Wald chi2(5)    =      80.83
Log simulated-pseudolikelihood = -197.16733       Prob > chi2     =     0.0000

                                    (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
travelmode   |
  travelcost |  -.0081881   .0035278    -2.32   0.020    -.0151024   -.0012739
    termtime |  -.0577873   .0089022    -6.49   0.000    -.0752353   -.0403393
-------------+----------------------------------------------------------------
air          |  (base alternative)
-------------+----------------------------------------------------------------
train        |
      income |  -.0378831   .0106206    -3.57   0.000     -.058699   -.0170671
       _cons |   .0389468   .4520208     0.09   0.931    -.8469978    .9248913
-------------+----------------------------------------------------------------
bus          |
      income |  -.0159619    .009725    -1.64   0.101    -.0350226    .0030988
       _cons |  -1.113611   .4240609    -2.63   0.009    -1.944755   -.2824666
-------------+----------------------------------------------------------------
car          |
      income |   .0046124   .0075777     0.61   0.543    -.0102397    .0194645
       _cons |  -3.713488   .6062384    -6.13   0.000    -4.901693   -2.525283
------------------------------------------------------------------------------
```

Example 2: Determinants of travel mode choice (VIII)

```
asmprobit choice travelcost termtime, case(id) alternatives(travelmode) casevars(income) stddev(homoskedastic) intmethod(halton)
intpoints(1000)robust

Alternative-specific multinomial probit        Number of obs      =        840
Case variable: id                              Number of cases    =        210

Alternative variable: travelmode               Alts per case: min =          4
                                                              avg =        4.0
                                                              max =          4

Integration sequence:                Halton
Integration points:                    1000       Wald chi2(5)    =      44.68
Log simulated-pseudolikelihood = -190.32744       Prob > chi2     =     0.0000

                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
travelmode   |
  travelcost |  -.0106986    .002702    -3.96   0.000    -.0159945   -.0054028
    termtime |  -.0398616   .0153844    -2.59   0.010    -.0700145   -.0097087
-------------+----------------------------------------------------------------
air          |  (base alternative)
-------------+----------------------------------------------------------------
train        |
      income |  -.0295295   .0102082    -2.89   0.004    -.0495373   -.0095217
       _cons |   .6141844   .5371857     1.14   0.253    -.4386803    1.667049
-------------+----------------------------------------------------------------
bus          |
      income |  -.0128889    .008084    -1.59   0.111    -.0287332    .0029554
       _cons |  -.0709518   .6799754    -0.10   0.917    -1.403679    1.261775
-------------+----------------------------------------------------------------
car          |
      income |  -.0045616    .008916    -0.51   0.609    -.0220367    .0129134
       _cons |  -1.976427   1.265471    -1.56   0.118    -4.456705    .5038517
-------------+----------------------------------------------------------------
   /atanhrP1 |   1.010803   .4863122     2.08   0.038     .0576485    1.963957
   /atanhrP2 |   .6224803   .4501503     1.38   0.167    -.2597982    1.504759
   /atanhrP3 |   1.126551   .3673321     3.07   0.002     .4065934    1.846509
-------------+----------------------------------------------------------------
      rho3_2 |   .7660939   .2008956                      .0575847    .9613907
      rho4_2 |   .5528526   .3125637                     -.2541067    .9060045
      rho4_3 |   .8098356   .1264234                      .3855763    .9514161
------------------------------------------------------------------------------
```

Example 2: Determinants of travel mode choice (IX)

```
asmprobit choice travelcost termtime, case(id) alternatives(travelmode) casevars(income)

Alternative-specific multinomial probit        Number of obs      =        840
Case variable: id                              Number of cases    =        210

Alternative variable: travelmode               Alts per case: min =          4
                                                              avg =        4.0
                                                              max =          4

Integration sequence:      Hammersley
Integration points:               200             Wald chi2(5)    =      32.05
Log simulated-likelihood = -190.09418             Prob > chi2     =     0.0000

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
travelmode   |
  travelcost |    -.00977   .0027834    -3.51   0.000    -.0152253   -.0043146
    termtime |  -.0377095   .0094088    -4.01   0.000    -.0561504   -.0192686
-------------+----------------------------------------------------------------
air          |  (base alternative)
-------------+----------------------------------------------------------------
train        |
      income |  -.0291971   .0089246    -3.27   0.001     -.046689   -.0117052
       _cons |   .5616376   .3946551     1.42   0.155    -.2118721    1.335147
-------------+----------------------------------------------------------------
bus          |
      income |  -.0127503   .0079267    -1.61   0.108    -.0282863    .0027857
       _cons |  -.0571364   .4791861    -0.12   0.905    -.9963239     .882051
-------------+----------------------------------------------------------------
car          |
      income |  -.0049086   .0077486    -0.63   0.526    -.0200957    .0102784
       _cons |  -1.833393   .8186156    -2.24   0.025     -3.43785   -.2289357
-------------+----------------------------------------------------------------
     /lnl2_2 |  -.5502039   .3905204    -1.41   0.159     -1.31561    .2152021
     /lnl3_3 |  -.6005552   .3353292    -1.79   0.073    -1.257788    .0566779
-------------+----------------------------------------------------------------
       /l2_1 |   1.131518   .2124817     5.33   0.000     .7150612    1.547974
       /l3_1 |   .9720669   .2352116     4.13   0.000     .5110606    1.433073
       /l3_2 |   .5197214   .2861552     1.82   0.069    -.0411325    1.080575
------------------------------------------------------------------------------
```

Example 2: Determinants of travel mode choice (X)

```
asmprobit choice travelcost termtime, case(id) alternatives(travelmode) casevars(income) structural

Alternative-specific multinomial probit        Number of obs      =        840
Case variable: id                              Number of cases    =        210

Alternative variable: travelmode               Alts per case: min =          4
                                                              avg =        4.0
                                                              max =          4

Integration sequence:      Hammersley
Integration points:               200             Wald chi2(5)    =      32.05
Log simulated-likelihood = -190.09418             Prob > chi2     =     0.0000

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
travelmode   |
  travelcost |  -.0097703   .0027834    -3.51   0.000    -.0152257   -.0043149
    termtime |  -.0377103   .0094092    -4.01   0.000     -.056152   -.0192687
-------------+----------------------------------------------------------------
air          |  (base alternative)
-------------+----------------------------------------------------------------
train        |
      income |  -.0291975   .0089246    -3.27   0.001    -.0466895   -.0117055
       _cons |   .5616448   .3946529     1.42   0.155    -.2118607     1.33515
-------------+----------------------------------------------------------------
bus          |
      income |    -.01275   .0079266    -1.61   0.108    -.0282858    .0027858
       _cons |  -.0571664   .4791996    -0.12   0.905    -.9963803    .8820476
-------------+----------------------------------------------------------------
car          |
      income |  -.0049085   .0077486    -0.63   0.526    -.0200955    .0102785
       _cons |  -1.833444   .8186343    -2.24   0.025    -3.437938     -.22895
-------------+----------------------------------------------------------------
   /lnsigma3 |  -.2447428   .4953363    -0.49   0.621    -1.215584    .7260985
   /lnsigma4 |  -.3309429   .6494493    -0.51   0.610     -1.60384    .9419543
-------------+----------------------------------------------------------------
  /atanhr3_2 |    1.01193   .3890994     2.60   0.009      .249309    1.774551
  /atanhr4_2 |   .5786576   .3940461     1.47   0.142    -.1936586    1.350974
  /atanhr4_3 |   .8885204   .5600561     1.59   0.113    -.2091693     1.98621
-------------+----------------------------------------------------------------
      sigma1 |          1  (base alternative)
      sigma2 |          1  (scale alternative)
      sigma3 |   .7829059   .3878017                      .2965368       2.067
      sigma4 |   .7182462   .4664645                      .2011227    2.564989
-------------+----------------------------------------------------------------
      rho3_2 |    .766559   .1604596                       .244269    .9441061
      rho4_2 |   .5216891   .2868027                     -.1912734     .874283
      rho4_3 |   .7106622    .277205                     -.2061713    .9630403
------------------------------------------------------------------------------
```

Example 2: Determinants of travel mode choice (XI)

Estimated WTP (i.e. US dollars the persons are on average approximately willing to pay for a one minute less terminal time):

- $\widehat{WTP}_{termtime}$ = -(-0.058/(-0.008)) = -7.06 in the first approach
- $\widehat{WTP}_{termtime}$ = -(-0.040/(-0.011)) = -3.73 in the second approach
- $\widehat{WTP}_{termtime}$ = -(-0.038/(-0.010)) = -3.86 in the last two approaches

Simulated Wald and likelihood ratio tests of the null hypothesis that neither travelcost nor income has an effect on the travel mode choice (on the basis of the third SML estimation):

```
asmprobit choice travelcost termtime, case(id) alternatives(travelmode) casevars(income)

estimates store unrestricted

test travelcost income

 ( 1)  [travelmode]travelcost = 0
 ( 2)  [train]income = 0
 ( 3)  [bus]income = 0
 ( 4)  [car]income = 0

           chi2(  4) =   26.52
         Prob > chi2 =    0.0000

asmprobit choice termtime, case(id) alternatives(travelmode)

estimates store restricted

lrtest unrestricted restricted

Likelihood-ratio test                                 LR chi2(4)  =     46.81
(Assumption: restricted nested in unrestricted)       Prob > chi2 =    0.0000
```

Example 2: Determinants of travel mode choice (XII)

Simulated likelihood ratio tests of the null hypothesis of the independent multinomial probit model versus the alternative hypotheses of the multinomial probit model that only restricts the variances of ∑ to the value one and of the most flexible multinomial probit model as well as the null hypothesis of the restricted variances model versus the alternative hypothesis of the most flexible multinomial probit model (with robustly estimated variance, respectively):

```
asmprobit choice travelcost termtime, case(id) alternatives(travelmode) casevars(income)
correlation(independent) stddev(homoskedastic) robust

estimates store independent

asmprobit choice travelcost termtime, case(id) alternatives(travelmode) casevars(income)
stddev(homoskedastic) robust

estimates store varrestricted

asmprobit choice travelcost termtime, case(id) alternatives(travelmode) casevars(income) robust

estimates store mostflexible

lrtest varrestricted independent, force

Likelihood-ratio test                                 LR chi2(3)  =     13.68
(Assumption: independent nested in varrestricted)     Prob > chi2 =    0.0034

lrtest mostflexible independent, force

Likelihood-ratio test                                 LR chi2(5)  =     14.15
(Assumption: independent nested in mostflexible)      Prob > chi2 =    0.0147

lrtest mostflexible varrestricted, force

Likelihood-ratio test                                 LR chi2(2)  =      0.47
(Assumption: varrestricted nested in mostflexible)    Prob > chi2 =    0.7917
```

Example 2: Determinants of travel mode choice (XIII)

Estimation of marginal probability effects for travelcost and income at the means of the explanatory variables (on the basis of the third SML estimation):

```
estat mfx, varlist(travelcost income)

Pr(choice = air) = .29434926
-------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+-----------------------------------------------------------------
travelcost   |
         air | -.002688   .000677   -3.97   0.000  -.004015  -.001362    102.65
       train |    .0009   .000436    2.07   0.039   .000046   .001755     130.2
         bus |  .000376   .000271    1.39   0.166  -.000155   .000908    115.26
         car |  .001412    .00051    2.77   0.006   .000412   .002412    95.414
-------------+-----------------------------------------------------------------
casevars     |
      income |  .003891   .001847    2.11   0.035   .000271   .007511    34.548
-------------------------------------------------------------------------------

Pr(choice = train) = .29531182
-------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+-----------------------------------------------------------------
travelcost   |
         air |  .000899   .000436    2.06   0.039   .000045   .001753    102.65
       train | -.004081   .001466   -2.78   0.005  -.006953  -.001208     130.2
         bus |  .001278    .00063    2.03   0.042   .000043   .002513    115.26
         car |  .001904   .000887    2.15   0.032   .000166   .003641    95.414
-------------+-----------------------------------------------------------------
casevars     |
      income |  -.00957   .002223   -4.31   0.000  -.013927  -.005214    34.548
-------------------------------------------------------------------------------
```

Example 2: Determinants of travel mode choice (XIV)

Pr(choice = bus) = .08880039

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
travelcost							
air	.00038	.000274	1.39	0.165	-.000157	.000916	102.65
train	.001279	.00063	2.03	0.042	.000044	.002514	130.2
bus	-.003182	.001175	-2.71	0.007	-.005485	-.00088	115.26
car	.001523	.000675	2.26	0.024	.0002	.002847	95.414
casevars							
income	.000435	.001461	0.30	0.766	-.002428	.003298	34.548

Pr(choice = car) = .32168607

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
travelcost							
air	.00141	.000509	2.77	0.006	.000411	.002408	102.65
train	.001903	.000886	2.15	0.032	.000166	.003641	130.2
bus	.001523	.000675	2.25	0.024	.000199	.002847	115.26
car	-.004836	.001539	-3.14	0.002	-.007853	-.001819	95.414
casevars							
income	.005246	.002166	2.42	0.015	.001002	.00949	34.548

Example 2: Determinants of travel mode choice (XV)

Now, SML estimations of the mixed logit model are considered. As discussed before, this requires the formation of three interaction terms (i.e. train_income, bus_income, car_income) if income is included as explanatory variable. With respect to the estimation of WTP, the parameter of travelcost is assumed to be fixed (besides the parameters of the interaction terms and the alternative specific constants), whereas the parameter of termtime is assumed to be random. The following estimations and tests (on the basis of the second SML estimation) are considered:

- Standard SML estimation without robustly estimated variances and SML estimation with R = 1000 Halton draws and robustly estimated variances
- Simulated Wald and likelihood ratio tests of the null hypothesis that neither travelcost nor income has an effect on the travel mode choice
- Simulated Wald test of the null hypothesis that the mean and standard deviation of the parameter of termtime are zero
- Simulated likelihood ratio test of the null hypothesis that the multinomial logit model is valid (which is in fact a mixture of a simulated and unsimulated likelihood ratio test)

The corresponding Stata commands lead to the following results:

Example 2: Determinants of travel mode choice (XVI)

```
mixlogit choice travelcost income_train income_bus income_car train bus car, group(id)
rand(termtime)

Mixed logit model                                 Number of obs   =        840
                                                  LR chi2(1)      =      29.82
Log likelihood = -174.61738                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Mean         |
  travelcost |  -.0194993   .0079791    -2.44   0.015     -.035138   -.0038605
income_train |  -.0717241    .022209    -3.23   0.001     -.115253   -.0281952
  income_bus |  -.0339118   .0227386    -1.49   0.136    -.0784787     .010655
  income_car |   .0076788   .0327348     0.23   0.815    -.0564802    .0718379
       train |   .3827847   .7554399     0.51   0.612     -1.09785     1.86342
         bus |  -1.544218   .8939996    -1.73   0.084    -3.296425    .2079885
         car |  -10.97316   2.519355    -4.36   0.000      -15.911    -6.03531
    termtime |  -.1914373   .0407844    -4.69   0.000    -.2713733   -.1115013
-------------+----------------------------------------------------------------
SD           |
    termtime |   .1154228   .0358206     3.22   0.001     .0452156    .1856299
------------------------------------------------------------------------------
```

→ $\widehat{WTP}_{termtime} = -(-0.191 / (-0.019)) = -9.82$

Example 2: Determinants of travel mode choice (XVII)

```
mixlogit choice travelcost income_train income_bus income_car train bus car, group(id)
rand(termtime) nrep(1000) robust

Mixed logit model                                  Number of obs   =        840
                                                   Wald chi2(8)    =      56.87
Log likelihood = -174.24838                        Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |               Robust
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Mean         |
  travelcost |  -.0198103   .0076932    -2.58   0.010    -.0348886   -.004732
income_train |   -.070943   .0263092    -2.70   0.007    -.1225081   -.0193779
  income_bus |  -.0337096   .0226769    -1.49   0.137    -.0781556    .0107364
  income_car |    .007999    .026458     0.30   0.762    -.0438578    .0598557
       train |   .3239961   .9328153     0.35   0.728    -1.504288     2.15228
         bus |  -1.590185    .919146    -1.73   0.084    -3.391678    .2113082
         car |  -10.97106   1.823578    -6.02   0.000    -14.54521   -7.396912
    termtime |  -.1917633   .0309725    -6.19   0.000    -.2524682   -.1310584
-------------+----------------------------------------------------------------
SD           |
    termtime |   .1131678   .0309098     3.66   0.000     .0525857    .1737499
------------------------------------------------------------------------------

estimates store mixedlogitunrestricted
```

→ $\widehat{WTP}_{termtime}$ = -(-0.192/(-0.020)) = -9.68

Example 2: Determinants of travel mode choice (XVIII)

```
test travelcost income_train income_bus income_car

 ( 1)  [Mean]travelcost = 0
 ( 2)  [Mean]income_train = 0
 ( 3)  [Mean]income_bus = 0
 ( 4)  [Mean]income_car = 0

           chi2(  4) =   18.35
         Prob > chi2 =    0.0011

test termtime

 ( 1)  [Mean]termtime = 0
 ( 2)  [SD]termtime = 0

           chi2(  2) =   40.80
         Prob > chi2 =    0.0000

mixlogit choice train bus car, group(id) rand(termtime) nrep(1000) robust

estimates store mixedlogitrestricted

lrtest mixedlogitunrestricted mixedlogitrestricted, force

Likelihood-ratio test                                 LR chi2(4)  =     34.57
(Assumption: mixedlogitre~d nested in mixedlogitun~d) Prob > chi2 =    0.0000

asclogit choice travelcost termtime, case(id) alternatives(travelmode) casevars(income)
robust

estimates store multlogitunrestricted

lrtest mixedlogitunrestricted multlogitunrestricted, force

Likelihood-ratio test                                 LR chi2(1)  =     30.55
(Assumption: multlogitunr~d nested in mixedlogitun~d) Prob > chi2 =    0.0000
```

Example 2: Determinants of travel mode choice (XIX)

Interpretation:

- The estimation results for the explanatory variables are qualitatively very similar to the former results
- The first two null hypotheses can be rejected at very low significance levels on the basis of simulated Wald tests
- The simulated likelihood ratio test statistic for the null hypothesis that neither travelcost nor income has an effect is 34.57 so that (in line with the simulated Wald test) the hypothesis can be rejected at all common significance levels
- The (simulated) likelihood ratio test statistic for the null hypothesis that the multinomial logit model is valid is 30.55 so that the hypothesis can be rejected at all common significance levels. This in line with the estimated standard deviation of the parameter of termtime, which is highly significantly different from zero.

Example 2: Determinants of travel mode choice (XX)

Finally, the ML estimation of a latent class logit model with two classes is considered. The “lclogit” command requires the same data organization as in the case of the “asclogit” and “asmprobit” commands. The following explanatory variables for the choice among the travel modes are included:

- Travelcost and termtime as alternative specific attributes
- Income as individual characteristic, which also has to be interacted with each alternative specific constant except for the base alternative air (income_train, income_bus, income_car)

In addition, income is included in the class membership model as explanatory variable.

While the “lclogit” command leads to the ML estimation on the basis of the EM algorithm, it does not generate estimations of the variance covariance matrix of the estimated parameters. This can be done with the postestimation command “lclogitml”, which uses the Newton-Raphson algorithm for the maximum likelihood estimation. The corresponding Stata command for two classes leads to the following results:

Example 2: Determinants of travel mode choice (XXI)

```
lclogit choice travelcost termtime income_train income_bus income_car, group(id) id(id)
membership(income) nclasses(2)

Latent class model with 2 latent classes

Choice model parameters and average classs shares
-------------------------------
    Variable |  Class1   Class2
-------------+-----------------
  travelcost |  -0.194    0.029
    termtime |  -0.025   -0.033
income_train |   0.303   -0.078
  income_bus |   0.266   -0.087
  income_car |   0.182   -0.095
-------------+-----------------
 Class Share |   0.499    0.501
-------------------------------

Class membership model parameters : Class2 = Reference class
-------------------------------
    Variable |  Class1   Class2
-------------+-----------------
      income |   0.009    0.000
       _cons |  -0.307    0.000
-------------------------------

Note: Model estimated via EM algorithm
```

Example 2: Determinants of travel mode choice (XXII)

```
lclogitml

Latent class model with 2 latent classes
------------------------------------------------------------------------------
      choice | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
choice1      |
  travelcost |  -.1937119   .0570851    -3.39   0.001    -.3055965   -.0818272
    termtime |  -.0252196   .0351362    -0.72   0.473    -.0940853    .0436461
income_train |   .3026577   .0986104     3.07   0.002      .109385    .4959304
  income_bus |   .2655774   .0926799     2.87   0.004     .0839282    .4472267
  income_car |   .1821022   .0842171     2.16   0.031     .0170397    .3471648
-------------+----------------------------------------------------------------
choice2      |
  travelcost |   .0287238   .0111881     2.57   0.010     .0067954    .0506521
    termtime |  -.0326176    .010148    -3.21   0.001    -.0525074   -.0127279
income_train |  -.0783855   .0227571    -3.44   0.001    -.1229886   -.0337824
  income_bus |  -.0867567   .0217786    -3.98   0.000     -.129442   -.0440714
  income_car |   -.094527   .0271546    -3.48   0.000    -.1477491    -.041305
-------------+----------------------------------------------------------------
share1       |
      income |   .0087331   .0095227     0.92   0.359    -.0099311    .0273973
       _cons |  -.3065039   .4689535    -0.65   0.513    -1.225636    .6126281
------------------------------------------------------------------------------
```

Example 2: Determinants of travel mode choice (XXIII)

```
lclogitml, iterate(15)

Latent class model with 2 latent classes
------------------------------------------------------------------------------
      choice | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
choice1      |
  travelcost |  -.1941678   .0588436    -3.30   0.001    -.3094992   -.0788364
    termtime |  -.0253871   .0359268    -0.71   0.480    -.0958023    .0450281
income_train |   .3032406   .1003832     3.02   0.003     .1064931     .499988
  income_bus |   .2661913   .0945176     2.82   0.005     .0809402    .4514425
  income_car |   .1823963   .0851249     2.14   0.032     .0155545     .349238
-------------+----------------------------------------------------------------
choice2      |
  travelcost |   .0286392    .011309     2.53   0.011      .006474    .0508044
    termtime |  -.0326248   .0101655    -3.21   0.001    -.0525488   -.0127008
income_train |  -.0782074   .0229795    -3.40   0.001    -.1232465   -.0331684
  income_bus |  -.0866666   .0220294    -3.93   0.000    -.1298433   -.0434899
  income_car |  -.0944761    .027268    -3.46   0.001    -.1479204   -.0410318
-------------+----------------------------------------------------------------
share1       |
      income |   .0087953   .0097396     0.90   0.367     -.010294    .0278845
       _cons |   -.310805    .481352    -0.65   0.518    -1.254238    .6326276
------------------------------------------------------------------------------
```

Example 2: Determinants of travel mode choice (XXIV)

Interpretation:

- While “id()” directly refers to the identification of the persons, “group()” refers to the identification of the choice sets in stated choice analyses (i.e. to different observations per individual). In this case with only one observation per person, the same name (here “id”) must be used for both required options. Furthermore, “choice” is a possible name for the dependent variable.
- With respect to the maximum number of iterations in the ML estimation with the EM algorithm, the default is 150. The number can be individually determined by including the additional command “iterate(#)”. Furthermore, the tolerance level for the increase of the log-likelihood function in the last five iterations before reaching the maximum value can be determined by the additional command “convergence(#)”, where the default is 0.00001.
- The command “lclogitml” after the ML estimation with the EM algorithm considers the log-likelihood function and the scores at the “lclogit” estimates
- In contrast, the additional inclusion of the command “iterate(#)” refers to a hybrid estimation strategy since it specifies the maximum number of iterations in the additional ML estimation with the Newton-Raphson algorithm on the basis of the “lclogit” estimates before. The default for this number is zero with no second ML estimation.

Example 2: Determinants of travel mode choice (XXV)

- Income has no significant effect on the membership in class 1
- Travelcost for a specific travel mode has a significantly negative effect on the choice of the same travel mode in class 1, but a significantly positive effect in class 2
- Termtime for a specific travel mode has a significantly negative effect on the choice of the same travel mode in class 2
- Income has a significantly positive effect on the (probability of the) choice of train, bus, and car compared with air in class 1
- Income has a significantly negative effect on the (probability of the) choice of train, bus, and car compared with air in class 2
- However, due to the strong instability of the estimation results with different seed values, latent class logit model estimations should not be considered in this example

---