Silberhorn, Nadja; Boztuğ, Yasemin; Hildebrandt, Lutz

Working Paper

Estimation with the nested logit model: specifications and software particularities

SFB 649 Discussion Paper, No. 2006,017

Provided in Cooperation with:
Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

Suggested Citation: Silberhorn, Nadja; Boztuğ, Yasemin; Hildebrandt, Lutz (2006) : Estimation with the nested logit model: specifications and software particularities, SFB 649 Discussion Paper, No. 2006,017, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:
https://hdl.handle.net/10419/25100

SFB 649 Discussion Paper 2006-017

Estimation with the Nested Logit Model: Specifications and Software Particularities

Nadja Silberhorn*
Yasemin Boztuğ*
Lutz Hildebrandt*

* Institute of Marketing, Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Estimation with the Nested Logit Model: Specifications and Software Particularities[1]

Nadja Silberhorn, Yasemin Boztuğ, Lutz Hildebrandt
Institute of Marketing, Humboldt University Berlin
Spandauer Straße 1, 10099 Berlin
{silberhorn,boztug,hildebr}@wiwi.hu-berlin.de

Abstract

Due to its ability to allow and account for similarities between pairs of alternatives, the nested logit model is increasingly used in practical applications. However the fact that there are two different specifications of the nested logit model has not received adequate attention. The *utility maximization nested logit* (UMNL) model and the *non-normalized nested logit* (NNNL) model have different properties, influencing the estimation results in a different manner. As the NNNL specification is not consistent with *random utility theory* (RUT), the UMNL form is preferred. This article introduces distinct specifications of the nested logit model and indicates particularities arising from model estimation. Additionally, it demonstrates the performance of simulation studies with the nested logit model. In simulation studies with the nested logit model using NNNL software (e. g. *PROC MDC* in SAS©), it must be pointed out that the simulation of the utility function's error terms needs to assume RUT-conformity. But as the NNNL specification is not consistent with RUT, the input parameters cannot be reproduced without imposing restrictions. The effects of using various software packages on the estimation results of a nested logit model are shown on the basis of a simulation study.

Keywords: nested logit model, utility maximization nested logit, non-normalized nested logit, simulation study

JEL-Codes: C13, C51, C87, M31

[1]Financial support by the Deutsche Forschungsgemeinschaft (DFG) through the research project #BO1952/1 and the SFB 649 "Economic Risk" is gratefully acknowledged.

1 Introduction

Modelling discrete choice decisions in the context of random utility theory is usually done with the multinomial logit model (MNL) (Guadagni and Little, 1983). But the MNL assumes proportional substitution patterns (Independence of Irrelevant Alternatives, IIA). To overcome this restrictive assumption, the nested logit model can be used for estimation in practical applications (Guadagni and Little, 1998; de Dios Ortúzar, 2001). The nested logit model admits more general substitution patterns and nevertheless remains, in contrast to the probit model for example, analytically tractable.
The existence of two unequal forms of the nested logit model has been underresearched so far. The *utility maximization nested logit* (UMNL) model and the *non-normalized nested logit* (NNNL) model have different properties which impact the estimation results. In many publications, the specification used is not explicitly mentioned. Both in simulation studies and in model estimations with real data, the implemented nested logit model specification within the software needs to be considered.
If there are only alternative-specific coefficients in the model, the software and thus the nested logit specification chosen can be accommodated merely by a nest-specific re-scaling of the estimated coefficients obtained from the NNNL software before interpretation. As soon as a generic coefficient enters the model, the *non-normalized nested logit* model is not consistent with random utility theory without imposing restrictions on the scale parameters.
Section 2 introduces the nested logit model and its application in marketing. In Section 3.1 the nested logit model is presented in general, whereas Section 3.2 introduces the two different forms of the nested logit model. In Section 3.3 their consistency with random utility theory is revised. Section 3.4 deals with the simulation of error terms and the necessary assumptions on their distribution. Section 4 goes into detail regarding the particularities in model estimation with NNNL software. This addressed difficulty is clarified with a simulation study in Section 5. Section 6 concludes with a summary.

2 Discrete Choice Models

Utility-based choice or choice based on the relative attractiveness of competing alternatives from a set of mutually exclusive alternatives is called a *discrete choice* situation. *Discrete choice* models are interpreted in terms of

an underlying behavioral model, the so called *random utility maximization (RUM)* model. The decision-maker chooses the alternative with the highest utility. Characteristics of the decision-maker and of the choice alternatives determine the alternatives' utilities. Demographics do not have a direct utility contribution *per se*, but serve as proxies for observable consumer heterogeneity.

Modelling discrete consumer decisions is characterized by a trade-off between flexibility and ease of the estimation (Munizaga and Alvarez-Daziano, 2001). On the one hand, there are logit models which are distinguished by closed choice probabilities but, due to restrictive substitution patterns, are often not very realistic. On the other hand, probit models assume a correlation structure of the error terms, the estimation of which can become very complex because of multidimensional integrals.

Discrete choice decisions in the context of random utility theory are usually modelled and estimated with the multinomial logit model (MNL) (Guadagni and Little, 1983). But the MNL assumes proportional substitution patterns (IIA), i. e. the ratio of the choice probabilities of two alternatives is not dependent on the presence or absence of other alternatives in the model. To overcome these restrictive substitution assumptions between alternatives, various extensions of the MNL exist, all with the general solution of allowing correlations between the alternatives' error terms. The idea of the nested logit model lies therefore in the grouping of similar alternatives into nests and thus creating a hierarchical structure of the alternatives (Ben-Akiva and Lerman, 1985; Train, 2003). The error terms of alternatives within a nest are correlated with each other, and the error terms of alternatives in different nests are uncorrelated.

Contrary to this, the mixed logit approach assumes that the stochastic influences are related to alternative-specific variables (Hensher and Greene, 2001). The error terms of similar alternatives are correlated. The stochastic part of the utility function is additively composed of an i.i.d. Gumbel-distributed term corresponding to the standard logit approach and a stochastic scaling parameter to account for heterogeneity. Not every single scaling parameter is estimated, but the parameters of their distribution instead.

The nested logit approach is predominantly used in the field of transportation research and logistics (Train, 1980; Bhat, 1997; Knapp et al, 2001), but can also be appropriate for marketing issues (Kannan and Wright, 1991; Chintagunta, 1993; Chintagunta and Vilcassim, 1998; Guadagni and Little, 1998; Chib et al, 2004). The nested logit model is the most often used hierarchical

model in marketing (Suárez et al, 2004) and can be used for modelling in any situation where subsets of alternatives share unobservable utility components (Ben-Akiva and Lerman, 1985). In the field of marketing the nested logit model is mainly applied in brand choice modelling (Kamakura et al, 1996; Ailawadi and Neslin, 1998; Guadagni and Little, 1998; Sun et al, 2003; Chib et al, 2004), where brands are nested, for example, regarding manufacturer (Anderson and de Palma, 1992); in a purchase incidence decision (Chintagunta, 1993; Chintagunta and Vilcassim, 1998); or regarding brand type (Baltas et al, 1997).
One important point to make is that the nested logit model is a combination of standard logit models. Marginal and conditional choice decisions are combined by a nesting structure (Hensher et al, 2005). The only goal of this process is to accommodate the violation of the IIA-assumption.
The nested logit model differs from the standard logit model in that the error components of the choice alternatives do not necessarily need to have the same distribution. Thus the nested logit model accounts for the fact that each alternative may have specific information in its unobservable utility component, which plays a role in the decision process. Subsets of alternatives may have similar information content, such that correlations between pairs of alternatives may exist (Hensher et al, 2005). The classification of alternatives regarding their similarities into nests and the thus resulting tree structure does not have anything in common with a stochastic valuation of alternatives within the scope of a decision tree. Nested logit models do not define the process of decision-finding, but account for differences in variances in the unobservable utility components (Hensher et al, 2005).

3 The Nested Logit Model

3.1 General Model Formulation

This article focuses on the example of a two-level nested logit model (see Figure 1). In this case, the choice probability P_{im} of an alternative i within nest m results from the product of the marginal choice probability P_m for nest m (Level 2) and the conditional choice probability $P_{i|m}$ for alternative i within nest m (Level 1). Both the marginal and the conditional probability are standard logit models. The *inclusive value* IV_m as the expected utility of nest m connects the two decision levels.

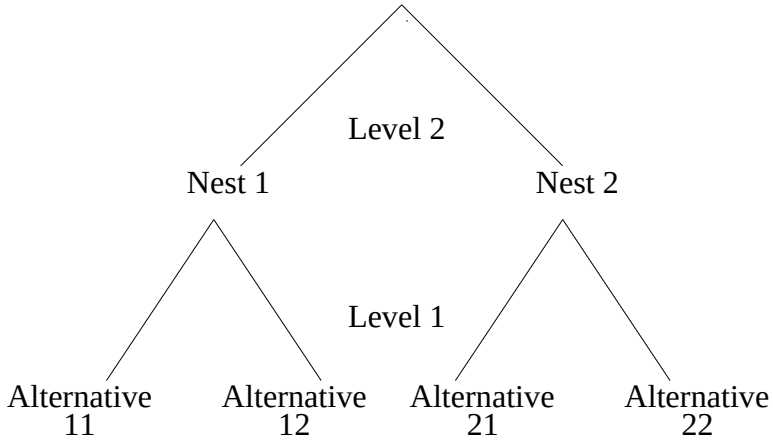

Figure 1: Tree structure of a nested logit model

The random utility U_{im} of alternative im results from the sum of a marginal utility component U_m from Level 2 and a conditional utility component $U_{i|m}$ from Level 1, which both consist of a deterministic part V and a stochastic part ν.

$$U_{im} = U_m + U_{i|m} = (V_m + \nu_m) + (V_{i|m} + \nu_{i|m}) \tag{1}$$

The error terms ν_m and $\nu_{i|m}$ are independent. The error terms $\nu_{i|m}$ are identically and independently distributed (i.i.d.) extreme-value with scale parameter μ_m. This can be interpreted as a measure of the correlation of the alternatives' errors within nest m (Heiss, 2002). The compound error terms ε_{im} are distributed such that the sum of U_m and $U^*_{i|m}$, the maximum of the $U_{i|m}$, is distributed extreme-value with scale parameter λ_m (Ben-Akiva and Lerman, 1985; Hunt, 2000).

$$Var(\nu_{i|m}) = \frac{\pi^2}{6\,\mu_m^2} \tag{2}$$

$$Var(\varepsilon_{im}) = Var(\nu_m + \nu^*_{i|m}) = \frac{\pi^2}{6\,\lambda_m^2}\,. \tag{3}$$

The scale parameters μ and λ describe the variances of the unobservable effects. Unconsidered utility components can variously impact the random components. This leads to different variances, which are explicitly accounted for by the introduction of these scale parameters. Each elemental alternative im has its own scale parameter $\mu^*_{i|m}$. But as these need to be equal

for all alternatives within a nest, the differentiation by i is redundant. The alternative-specific scale parameters $\mu^*_{i|m}$ are replaced by nest-specific scale parameters μ_m. The scale parameters λ_m are associated with the upper level, so that there is no need to replace them.
The compound unobservable utility components ε_{im} contain variance components both from the lower and the upper decision level. Thus the variances on the upper level cannot be smaller than those on the lower level. Therefore the scale parameters need to satisfy the following condition (Carrasco and de Dios Ortúzar, 2002; Hensher et al, 2005):

$$\lambda_m < \mu_m \,. \tag{4}$$

3.2 Different Nested Logit Model Specifications

Train (2003), Heiss (2002), Hunt (2000) and Koppelman and Wen (1998a,b) point to the existence of different nested logit model specifications and the issues arising from this regarding different estimation results.
The *non-normalized nested logit* (NNNL) model was derived from the standard logit model to relax the IIA-assumption. The elementary NNNL form is not consistent with utility maximization theory (Koppelman and Wen, 1998b). On the other hand, the *utility maximization nested logit* (UMNL) model, which was derived from McFadden's *Generalized Extreme Value* (GEV) theory (McFadden, 1978, 1981), is consistent with the utility maximization theory (Koppelman and Wen, 1998b).
The difference between these nested logit model specifications lies in the explicit scaling of the deterministic utility component in the UMNL form. In the case of generic coefficients, this means for the NNNL specification that the estimated parameters are indeed constant for all alternatives but not the hidden "true" parameters. The reason lies in the implicit nest-specific scaling within the NNNL specification (Heiss, 2002).
Table 1 compares the two specifications (Hunt, 2000; Koppelman and Wen, 1998a). The letters m and n represent the nests on Level 2, with $m \neq n$, and the letters i and j denote the elemental alternatives on Level 1, with $i \neq j$. The set of all elemental alternatives within nest m is called C_m.

Due to identification problems, one of the scale parameters in the *utility maximization nested logit* (UMNL) specification needs to be normalized to 1 (Daly, 2001; Hunt, 2000). A normalization on the lower Level 1 ($\mu_m = \mu_n = 1$) leads to the RU1 UMNL model; a normalization on the

Table 1: Specifications of the nested logit model

	UMNL *utility maximization* *nested logit*	NNNL *non-normalized* *nested logit*
P_m	$\frac{\exp(\lambda_m V_m + \frac{\lambda_m}{\mu_m} IV_m)}{\sum_n \exp(\lambda_n V_n + \frac{\lambda_n}{\mu_n} IV_n)}$	$\frac{\exp(V_m + \frac{1}{\mu_m} IV_m)}{\sum_n \exp(V_n + \frac{1}{\mu_n} IV_n)}$
$P_{i\|m}$	$\frac{\exp(\mu_m V_{i\|m})}{\sum_{j \in C_m} \exp(\mu_m V_{j\|m})}$	$\frac{\exp(V_{i\|m})}{\sum_{j \in C_m} \exp(V_{j\|m})}$
IV_m	$\ln \sum_{j \in C_m} \exp(\mu_m V_{j\|m})$	$\ln \sum_{j \in C_m} \exp(V_{j\|m})$

upper Level 2 ($\lambda_m = \lambda_n = 1$) results in the RU2 UMNL model.

3.3 Testing the Nested Logit Models Regarding Consistency with Random Utility Theory

To be consistent with utility maximization theory, each alternative's choice probability must not change when adding a constant term a to each alternative's deterministic utility component (Koppelman and Wen, 1998b).
Formally, this means that the new deterministic utility component $V^*_{i|m}$ results from the sum of the old deterministic utility component $V_{i|m}$ and a constant term a.

$$V^*_{i|m} = V_{i|m} + a \tag{5}$$

To be theory-consistent, the new choice probability (P^*_{im}) has to equal the old choice probability (P_{im}) for alternative im:

$$P^*_{im} = P_{im} \tag{6}$$

The procedure of testing for theory consistency is shown as an example with the *non-normalized nested logit* (NNNL) specification. The new marginal choice probability (P^*_m) is compared with the old marginal choice probability (P_m), the new conditional choice probability ($P^*_{i|m}$) is compared with the old conditional choice probability ($P_{i|m}$), and the new *inverse value* (IV^*_m) is

compared with the old *inverse value* (IV_m).

$$\begin{aligned} IV_m^* &= \ln \sum_{j \in C_m} \exp\left(V_{j|m} + a\right) \\ &= \ln \sum_{j \in C_m} \left(\exp\left(V_{j|m}\right) \exp\left(a\right)\right) \\ &= \ln \left(\exp\left(a\right) \sum_{j \in C_m} \exp\left(V_{j|m}\right)\right) \\ &= \ln\left(\exp\left(a\right)\right) + \ln \left(\sum_{j \in C_m} \exp\left(V_{j|m}\right)\right) \\ &= a + \ln \left(\sum_{j \in C_m} \exp\left(V_{j|m}\right)\right) \\ &= a + IV_m \end{aligned} \tag{7}$$

$$\begin{aligned} P_{i|m}^* &= \frac{\exp\left(V_{i|m} + a\right)}{\sum_{j \in C_m} \exp\left(V_{j|m} + a\right)} \\ &= \frac{\exp\left(V_{i|m}\right) \exp\left(a\right)}{\sum_{j \in C_m} \left(\exp\left(V_{j|m}\right) \exp\left(a\right)\right)} \\ &= \frac{\exp\left(V_{i|m}\right) \exp\left(a\right)}{\exp\left(a\right) \sum_{j \in C_m} \exp\left(V_{j|m}\right)} \\ &= \frac{\exp\left(V_{i|m}\right)}{\sum_{j \in C_m} \exp\left(V_{j|m}\right)} \\ &= P_{i|m} \end{aligned} \tag{8}$$

$$
\begin{aligned}
P_m^* &= \frac{\exp\left(V_m + \frac{1}{\mu_m} IV_m^*\right)}{\sum_n \exp\left(V_n + \frac{1}{\mu_n} IV_n^*\right)} \\
&= \frac{\exp\left(V_m + \frac{1}{\mu_m} (a + IV_m)\right)}{\sum_n \exp\left(V_n + \frac{1}{\mu_n} (a + IV_n)\right)} \\
&= \frac{\exp\left(V_m\right) \exp\left(\frac{a}{\mu_m}\right) \exp\left(\frac{1}{\mu_m} IV_m\right)}{\sum_n \exp\left(V_n\right) \exp\left(\frac{a}{\mu_n}\right) \exp\left(\frac{1}{\mu_n} IV_n\right)}
\end{aligned}
$$

only if $\mu_m = \mu_n = \mu$ holds, then

$$
\begin{aligned}
&= \frac{\exp\left(\frac{a}{\mu}\right) \exp\left(V_m\right) \exp\left(\frac{1}{\mu} IV_m\right)}{\exp\left(\frac{a}{\mu}\right) \sum_n \exp\left(V_n\right) \exp\left(\frac{1}{\mu} IV_n\right)} \\
&= \frac{\exp\left(V_m + \frac{1}{\mu} IV_m\right)}{\sum_n \exp\left(V_n + \frac{1}{\mu} IV_n\right)} \\
&= P_m \qquad (9)
\end{aligned}
$$

Analogous to this procedure, consistency with random utility theory can be tested for the Level 1 normalized ($\mu_m = \mu_n = 1$) *utility maximization nested logit* (RU1 UMNL) model and the Level 2 normalized ($\lambda_m = \lambda_n = 1$) *utility maximization nested logit* (RU2 UMNL) model.

Table 2 summarizes the results. In the NNNL and the RU1 UMNL specification, the new *inverse value* IV_m^* equals the sum of the old *inverse value* IV_m and the added constant term a. In the RU2 UMNL model, the added constant term a is additionally scaled with the scale parameter μ_m. While the new choice probability $P_{i|m}^*$ does not differ from the old choice probability $P_{i|m}$ in all three nested logit specifications, the new choice probability P_m^* on the upper level differs from the old one. Without imposing restrictions, only the RU2 UMNL specification satisfies the demand of consistency with utility theory. Only in the RU2 form does the choice probability P_{im} equal the choice probability P_{im}^* after adding a term a to the utility component $V_{i|m}$. In the RU1 UMNL specification, consistency can only be reached by imposing the restriction $\lambda_m = \lambda_n = \lambda$. As shown in (9), consistency with random utility theory can be ensured in the NNNL form by imposing the

Table 2: Nested logit specifications and utility maximization

	NNNL *non-normalized nested logit*	UMNL *utility maximization nested logit*	
		RU1 ($\mu_m = \mu_n = 1$)	RU2 ($\lambda_m = \lambda_n = 1$)
$V^*_{i\|m}$	$V_{i\|m} + a$	$V_{i\|m} + a$	$V_{i\|m} + a$
IV^*_m	$IV_m + a$	$IV_m + a$	$IV_m + a\,\mu_m$
$P^*_{i\|m}$	$P_{i\|m}$	$P_{i\|m}$	$P_{i\|m}$
P^*_m	$\neq P_m$	$\neq P_m$	P_m
P^*_{im}	$\neq P_{im}$	$\neq P_{im}$	P_{im}

restriction $\mu_m = \mu_n = \mu$.
The new choice probability of an alternative im results as the product of the new marginal choice probability P^*_m and the new conditional choice probability $P^*_{i|m}$. Because of the generally **not** theory-consistent results on the level of the marginal choice probabilities in the *non-normalized nested logit* (NNNL) and the Level 1 normalized *utility maximization nested logit* (RU1 UMNL) specification, only the Level 2 normalized *utility maximization nested logit* (RU2 UMNL) specification satisfies condition (6).

3.4 Simulation of Error Terms for a Nested Logit Model

A good and appropriate way to test model validity is to conduct a simulation study where the true parameters are known and correlations are determined. When the sample size is large, the estimated parameters should be very close to the true parameters (Cameron and Trivedi, 2005).
The nested logit model assumes error terms following an extreme-value type I (Gumbel) distribution with *location*-parameter η and *scale*-parameter ζ (Ben-Akiva and Lerman, 1985). This distribution has the density function

$$f(z) \;=\; \zeta * \exp\left[(\eta - z)\,\zeta\right] * \exp\left[-\exp\left[(\eta - z)\,\zeta\right]\right] \tag{10}$$

and the cumulative density function

$$F(z) \;=\; \exp\left[-\exp\left[(\eta - z)\,\zeta\right]\right]. \tag{11}$$

An *Extreme*[η;ζ]-distributed random variable Z has mean value

$$E(Z) \;=\; \eta + \frac{\gamma}{\zeta} \tag{12}$$

and variance

$$Var(Z) \;=\; \frac{\pi^2}{6\,\zeta^2} \tag{13}$$

with $\gamma = 0.577216$ denoting the *Euler* constant.
When simulating a utility function, mainly the stochastic part needs to be taken into account. In the nested logit model, the two components ν_m and $\nu_{i|m}$ are to be simulated and assumptions about their distributions are to be made. The variance of the error term $\nu_{i|m}$ can be distinguished from Equation (2). Further, the following is true (Hunt, 2000):

$$\eta_i \;=\; 0 \tag{14}$$

Combining the expression for the variance and (12) and (14) gives the mean value of the error term.

$$E(\nu_{i|m}) \;=\; \frac{\gamma}{\mu_m} \tag{15}$$

Besides assumptions on the error term $\nu_{i|m}$, assumptions on the error term ν_m are to be made as well. These do not explicitly arise from the model's assumptions but are to be derived separately. The variance of the error term $\nu^*_{i|m}$ equals the variance of the error term $\nu_{i|m}$ and also results from Equation (2). Hunt (2000) shows that

$$\eta^*_i \;=\; \frac{1}{\mu_m}\,\ln \sum_{i \in C_m} \exp\left(\mu_m\,\eta_i\right), \tag{16}$$

from which, together with (14), is yielded the following mean value:

$$E(\nu^*_{i|m}) \;=\; \frac{1}{\mu_m}\,\ln \sum_{i \in C_m} \exp\left(0\right) \;+\; \frac{\gamma}{\mu_m}\,. \tag{17}$$

The variance of the compound error term ε_{im} results from Equation (3), the mean value results according to the extreme-value distribution with $\eta^*_m = 0$ and $\zeta^*_m = \lambda_m$.

$$E(\varepsilon_{im}) \;=\; E(\nu_m \;+\; \nu^*_{i|m}) \;=\; \frac{\gamma}{\lambda_m} \tag{18}$$

After clarifying assumptions on the distribution of the error terms $\nu^*_{i|m}$ and ε_{im}, assumptions on the distribution of the error term ν_m can be derived.

$$\begin{aligned} E(\nu_m) &= E(\nu_m + \nu^*_{i|m}) - E(\nu^*_{i|m}) \\ &= -\left(\frac{1}{\mu_m} \ln \sum_{i \in C_m} \exp(0)\right) + \left(\frac{\gamma}{\left(\frac{\mu_m \lambda_m}{\mu_m - \lambda_m}\right)}\right) \end{aligned} \tag{19}$$

$$\begin{aligned} Var(\nu_m) &= Var(\nu_m + \nu^*_{i|m}) - Var(\nu^*_{i|m}) - 2\,Cov(\nu_m\, \nu^*_{i|m}) \\ &= \frac{\pi^2}{6\,\lambda_m^2} - \frac{\pi^2}{6\,\mu_m^2} - 0 \\ &= \frac{\pi^2}{6\left(\frac{\mu_m^2\,\lambda_m^2}{\mu_m^2 - \lambda_m^2}\right)} \end{aligned} \tag{20}$$

The error term $\nu_{i|m}$ (Level 1) follows an *Extreme*$[0; \mu_m]$-distribution, the error term ν_m (Level 2) is *Extreme*$\left[-\left(\frac{1}{\mu_m} \ln \sum_{i \in C_m} \exp(0)\right); \sqrt{\left(\frac{\mu_m^2\,\lambda_m^2}{\mu_m^2 - \lambda_m^2}\right)}\right]$-distributed. The simulation of a data set that is to be estimated with the nested logit model requires the simulation of deterministic utility components $V_{i|m}$ for Level 1 and V_m for Level 2 according to Equation (1), and of stochastic utility components $\nu_{i|m}$ and ν_m. Whereas the observable exogenous variables are relatively straightforward to simulate by imposing specific assumptions on the distribution and correlation patterns, the simulation of the unobservable influences requires falling back on the assumptions (2), (15), (19), and (20).

4 Estimation of Nested Logit Models

Before estimating a nested logit model with a specific software package, the implemented nested logit model specification (*utility maximization nested logit* or *non-normalized nested logit*) needs to be investigated.
The software packages **SAS**© (SAS, 2004) and **ALOGIT**© (see Carrasco and de Dios Ortúzar (2002)) use the *non-normalized nested logit* (NNNL) specification for model estimation. **STATA**© (Heiss, 2002), **GAUSS**© (Carrasco and de Dios Ortúzar, 2002) and **LIMDEP**© (Hunt, 2000; Hensher and

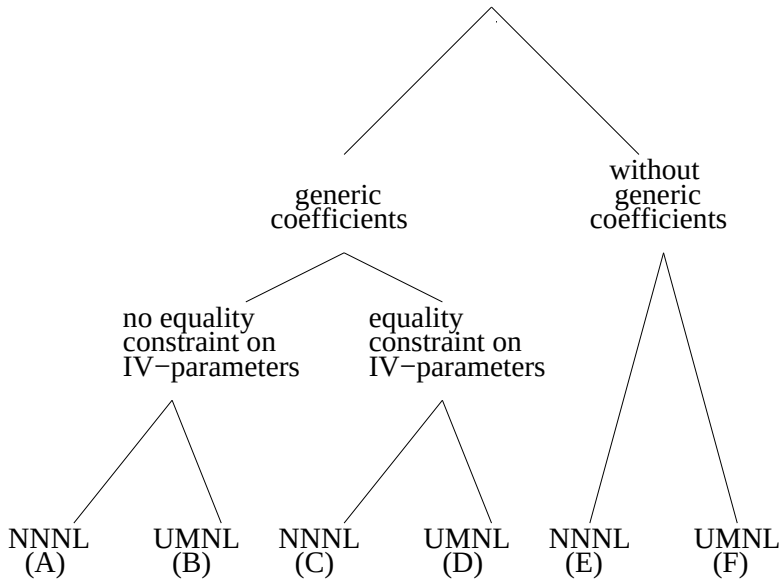

Figure 2: Overview model types

Greene, 2002) offer the possibility to choose between the *non-normalized nested logit* (NNNL) and the *utility maximization nested logit* (UMNL) specification.
In case only NNNL software is available, there are several particularities in model estimation to take into consideration. The crucial point is whether there are only alternative-specific coefficients in the model, or also at least one generic coefficient. Generic coefficients are constant for all alternatives. A variation on the utility contribution could be reached via alternative-specific values of the corresponding variables.
Moreover, Hunt (2000) points to the peculiarities of partially degenerate model structures. Nests with only one elemental alternative are called degenerate nests. For further and detailed information regarding the estimation procedure when degenerate nests enter the model, the reader is referred to the literature (Heiss, 2002; Hensher et al, 2005; Hunt, 2000).

4.1 Alternative-Specific Coefficients

If there are no generic coefficients in the model (models E and F in Figure 2), the *non-normalized nested logit* (NNNL) and the *utility maximization nested logit* (UMNL) specification are equivalent (Heiss, 2002). The coefficients esti-

mated with NNNL software are to be re-scaled with the according estimated IV-parameter. Only then is a correct interpretation possible. It must be taken into account which alternative belongs to which nest. The estimated alternative-specific coefficient $\beta_{i|m}$ has to be scaled with the corresponding nest-specific IV-parameter IV_m.

$$\beta_{i|m}^{\text{UMNL}} = IV_m^{\text{NNNL}} * \beta_{i|m}^{\text{NNNL}} \tag{21}$$

The models E and F are not focused on in detail, because in marketing models usually at least one variable with a generic coefficient, i. e. one exogenous variable with a constant coefficient for all alternatives, enters the model. Typically in modelling purchase decisions, this is the variable "price" as one of the central marketing-mix elements.

4.2 Generic Coefficients

Random utility maximizing models can generally not be estimated with *non-normalized nested logit* (NNNL) software when generic coefficients enter the model (model A in Figure 2).
As can be seen from Table 1, only in the *utility maximization nested logit* (UMNL) specification are the deterministic utility components V_m and $V_{i|m}$ scaled explicitly with the parameters λ_m and μ_m respectively. Table 3 refers to this with an example of the conditional deterministic utility component. The conditional deterministic utility component $V_{i|m}$ results as the product of a generic coefficient β and the alternative-specific values of the vector of the exogenous variables X_i.
Contrary to the explicit scaling in the UMNL specification, the coefficients

Table 3: Scaling of the deterministic utility component

NNNL *non-normalized* *nested logit*	UMNL *utility maximization* *nested logit*
$V_{i\|m} = \beta X_i$	$\mu_m V_{i\|m} = \mu_m \beta X_i$

in the NNNL specification are automatically and implicitly nest-specifically

scaled. The coefficients estimated in the NNNL model are thus not the "true" coefficients. In fact the estimated coefficients are constant for all alternatives, but not the hidden "true" coefficients. And this is a violation of the definition of generic coefficients.
By imposing restrictions it can be guaranteed that, even when using NNNL software, parameters consistent with random utility can be estimated (model C in Figure 2). It has to be assured that the coefficients in each nest are scaled equally. The *IV*-parameters are thus to be made equal for all nests. But, of course, each restriction on the parameter estimates means a loss of information in the data.
Studies have shown that the restricted form of the *non-normalized nested logit* (NNNL) model (model C in Figure 2) reproduces the estimation results of the restrictive Level 1 normalized *utility maximization nested logit* (RU1 UMNL) form (model D in Figure 2) (Heiss, 2002; Hensher and Greene, 2002; Hunt, 2000). Re-scaling the parameter estimates in the restrictive NNNL model with the estimated *IV*-parameter results in the parameter estimates of the restrictive Level 2 normalized *utility maximization nested logit* (RU2 UMNL) model.

$$\text{NNNL}_{res} = \text{RU1}_{res} \tag{22}$$

$$\text{NNNL}_{res} * IV_{\text{NNNL}_{res}} = \text{RU2}_{res} \tag{23}$$

Koppelman and Wen (1998a) have shown a second possibility to guarantee the consistency with utility maximizing theory without imposing restrictions on the *IV*-parameters. First, additional dummy nests below the lowest level are to be introduced into the model, and second, the thus additionally estimated scale parameters have to be defined in such a way that "the product of all the ratios of scale parameters between levels must be identical from the root to all elemental alternatives" (Hensher and Greene (2002), p. 13).

5 Simulation Study with a Software Comparison

As was shown in Section 3.3, without imposing restrictions, only the Level 2 normalized *utility maximization nested logit* (RU2 UMNL) specification is consistent with random utility theory. In the following, two simulated data sets (each having $n = 10,400$ observations) are generated with the software

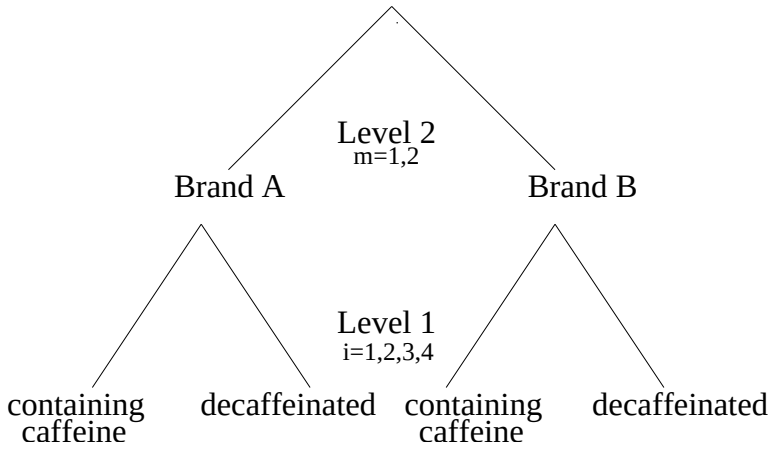

Figure 3: Two-level nested logit model

SAS© 9.1.3. When generating data set 1, a RU2 specification without restriction is assumed; generating data set 2, the restriction $\mu_m = \mu_n = \mu$ is imposed on a RU2 specification.
In this simulation study the coffee market is simulated in a very simplistic manner. The simulated market consists of only two brands A and B, where both offer variants containing caffeine and decaffeinated. Figure 3 shows the nest structure of this *discrete choice* situation.

According to Equation (1), the random utility U_{im} of each alternative im results from the sum of a marginal utility component U_m from Level 2 and a conditional utility component $U_{i|m}$ from Level 1, which both consist of a deterministic part V and a stochastic part ν. In this study, the deterministic marginal utility component V_m is neglected. It is often hard to find any variables that are nest- rather than alternative-specific. But even if a nest-specific variable does exist, specifying this variable for the nest or for all alternatives within this nest does not make a difference (Heiss, 2002). The stochastic marginal utility component ν_m, which captures all unobservable and omitted effects, must be integrated into the model despite the non-existence of the deterministic marginal utility component V_m. Consequently, the overall utility for this simulation study arises from

$$U_{im} = V_{i|m} + (\nu_{i|m} + \nu_m)\,. \tag{24}$$

Furthermore, the explanatory variables price (PRI), promotion (PRO), age of the decision maker (AGE) and sex of the decision maker (SEX) are included

in the model. Alternative-specific constants (ASC) are neglected in this simulation study, but must be integrated in the model when estimating with real data. The underlying deterministic conditional utility component for this simulation study is as follows

$$V_{i|m} = \phi_{i|m}\, AGE_h + \tau_{i|m}\, SEX_h + \beta_{pri}\, PRI_{i|m} + \beta_{pro}\, PRO_{i|m}\,. \qquad (25)$$

The variables PRI and PRO are such with generic coefficients (see section 4.2), i. e. they have a constant coefficient β for all alternatives. The alternative *containing caffein* in nest *Brand A* ($cc|A$) is declared as reference point, and its alternative-specific coefficients $\phi_{K|A}$ and $\tau_{K|A}$ are set to zero.
As was described in detail in Section 3.4, the error term $\nu_{i|m}$ on Level 1 is distributed extreme-value $[0; \mu_m]$, the error term ν_m on Level 2 is distributed extreme-value $\left[-\left(\frac{1}{\mu_m} \ln \sum_{i \in C_m} \exp(0)\right); \sqrt{\left(\frac{\mu_m^2\, \lambda_m^2}{\mu_m^2 - \lambda_m^2}\right)}\right]$.
For the simulation of the data sets 1 and 2, the following assumptions are made:

- age
  - AGE=1: *p=0.15*
  - AGE=2: *p=0.20*
  - AGE=3: *p=0.30*
  - AGE=4: *p=0.20*
  - AGE=5: *p=0.15*
- sex
  - female, SEX=0: *p=0.53*
  - male, SEX=1: *p=0.47*
- price
  - A containing caffein: *normal with* $[3.99; 0.20^2]$
  - A decaffeinated: *normal with* $[3.89; 0.20^2]$
  - B containing caffein: *normal with* $[4.29; 0.10^2]$
  - B decaffeinated: *normal with* $[4.19; 0.10^2]$

- promotion
  - A containing caffein: *uniform in [0;1], rounded to 0 or 1*
  - A decaffeinated: *uniform in [0;1], rounded to 0 or 1*
  - B containing caffein: *uniform in [0;1], rounded to 0 or 1*
  - B decaffeinated: *uniform in [0;1], rounded to 0 or 1*

Model estimation is done with the procedure *PROC MDC* in SAS© 9.1.3, and with the commands *nlogit* and *nlogitrum* in STATA© 9.1. The NNNL specification underlies the procedure *PROC MDC* and the command *nlogit* (see SAS (2004) and Heiss (2002)), and the RU2 UMNL specification underlies the command *nlogitrum* (see Heiss (2002)).

5.1 Models Without Equality Constraint on the *IV*-Parameters

According to the *utility maximization nested logit* (UMNL) RU2 specification, the scale parameters λ_A and λ_B are set equal to 1. When simulating data for the models A and B (Figure 2), the scale parameters μ_A and μ_B are **not** imposed by an equality constraint.
The coefficients of the exogenous variables generated with SAS© are estimated with SAS© and STATA©. Table 4 gives an overview of the simulated data sets for models A and B.
The data sets 1.1 to 1.4 share the fact that the NNNL model estimated

Table 4: Overview data sets for models A and B

	$\mu_A = 1.3$ $\mu_B = 1.7$	$\mu_A = 1.2$ $\mu_A = 1.5$
input-parameter set #1	**data set 1.1**	**data set 1.3**
input-parameter set #2	**data set 1.2**	**data set 1.4**

with the procedure *PROC MDC* in SAS© results in the same coefficients as

the NNNL model estimated with the command *nlogit* in STATA©. The RU2 UMNL model estimated with the command *nlogitrum* in STATA© results in different coefficients. The log-likelihood values and the *IV*-parameters $\left(\frac{\lambda}{\mu}\right)$ of the two *non-normalized nested logit* (NNNL) models are identical and differ from those of the RU2 UMNL model.
The *non-normalized nested logit* (NNNL) models estimated with *nlogit* in STATA© and *PROC MDC* in SAS© estimate an inverse relation of the *IV*-parameters in the two nests. The *IV*-parameter in nest A is smaller than the *IV*-parameter in nest B, resulting in a higher correlation $\left(1 - \left(\frac{\lambda}{\mu}\right)^2\right)$ of the alternatives within nest A than within nest B. When generating the simulated data, a higher correlation in nest B was assumed.
The Level 2 normalized *utility maximization nested logit* (UMNL RU2) model is consistent with random utility theory (see Table 2) even without imposing restrictions and should be able to reproduce the input-parameters with a high reliability. This can only partially be confirmed. Moreover, the estimated *IV*-parameter for the nest "Brand A" is expectedly higher than the one for the nest "Brand B", but the difference between those two *IV*-parameters is substantially smaller than it was assumed when simulating the data. At this point, one can not speak of a satisfying and reliable reproduction of the data. The reasons for this are unknown for the present but should be part of further investigations.
In the *utility maximization nested logit* (UMNL) model, the *IV*-parameters only capture the (dis-)similarity of the alternatives within the nest. The *IV*-parameters in the *non-normalized nested logit* (NNNL) model capture another effect: the relative importance of the variables with generic coefficients for the alternatives within the corresponding nest (see Heiss (2002), p. 240). Although these two effects are not in line, they are captured in the NNNL model with one single *IV*-parameter. The "generic" specification of the NNNL model implies a contradictory restriction. This is the reason why "generic" models should not be estimated with NNNL software without imposing restrictions.
Only if it is a priori assumed that the *IV*-parameters are the same in all nests, the scaling problem of the NNNL model can be avoided. The presence of generic coefficients then does not bias the estimates of the NNNL model, because the coefficients are equally scaled in each nest.

Table 5: Model estimation with data set 1.1 (RU2 without restriction)

parameter	input-parameter set **#1**	**SAS**© *PROC MDC* NNNL **(A)**	**STATA**© *nlogit* NNNL **(A)**	**STATA**© *nlogitrum* RU2 UMNL **(B)**
$\phi_{dc\|A}$	-0.50	-0.57***	-0.57***	-0.41***
$\phi_{cc\|B}$	0.50	0.60***	0.60***	0.21***
$\phi_{dc\|B}$	1.00	1.20***	1.20***	0.60***
$\tau_{dc\|A}$	2.00	2.29***	2.29***	1.60***
$\tau_{cc\|B}$	-1.00	-1.20***	-1.20***	-1.01***
$\tau_{dc\|B}$	-2.00	-2.46***	-2.46***	-1.79***
β_{pro}	3.00	3.63***	3.63***	2.42***
β_{pri}	-2.00	-2.25***	-2.25***	-1.97***
IV_A	0.77	0.55***	0.55***	0.75***
IV_B	0.59	0.74***	0.74***	0.64***
LL		-8,006	-8,006	-8,135

$n = 10,400$; *** $\alpha = 0.01$

Table 6: Model estimation with data set 1.2 (RU2 without restriction)

parameter	input-parameter set **#2**	**SAS**© *PROC MDC* NNNL **(A)**	**STATA**© *nlogit* NNNL **(A)**	**STATA**© *nlogitrum* RU2 UMNL **(B)**
$\phi_{dc\|A}$	-1.00	-1.20***	-1.20***	-0.84***
$\phi_{cc\|B}$	0.50	0.62***	0.62***	0.24***
$\phi_{dc\|B}$	0.50	0.62***	0.62***	0.23***
$\tau_{dc\|A}$	3.00	3.58***	3.58***	2.50***
$\tau_{cc\|B}$	-1.50	-1.78***	-1.78***	-1.28***
$\tau_{dc\|B}$	-1.00	-1.27***	-1.27***	-0.94***
β_{pro}	2.00	2.48***	2.48***	1.64***
β_{pri}	-1.50	-1.71***	-1.71***	-1.46***
IV_A	0.77	0.55***	0.55***	0.74***
IV_B	0.59	0.73***	0.73***	0.63***
LL		-9,551	-9,551	-9,640

$n = 10,400$; *** $\alpha = 0.01$

Table 7: Model estimation with data set 1.3 (RU2 without restriction)

parameter	input-parameter set **#1**	**SAS**© *PROC MDC* NNNL **(A)**	**STATA**© *nlogit* NNNL **(A)**	**STATA**© *nlogitrum* RU2 UMNL **(B)**
$\phi_{dc\|A}$	-0.50	-0.56***	-0.56***	-0.35***
$\phi_{cc\|B}$	0.50	0.60***	0.60***	0.22***
$\phi_{dc\|B}$	1.00	1.19***	1.19***	0.57***
$\tau_{dc\|A}$	2.00	2.27***	2.27***	1.40***
$\tau_{cc\|B}$	-1.00	-1.22***	-1.22***	-0.89***
$\tau_{dc\|B}$	-2.00	-2.44***	-2.44***	-1.58***
β_{pro}	3.00	3.58***	3.58***	2.14***
β_{pri}	-2.00	-2.19***	-2.19***	-1.59***
IV_A	0.83	0.51***	0.51***	0.65***
IV_B	0.67	0.64***	0.64***	0.58***
LL		-8,363	-8,363	-8,433

$n = 10,400$; *** $\alpha = 0.01$

Table 8: Model estimation with data set 1.4 (RU2 without restriction)

parameter	input-parameter set **#2**	**SAS**© *PROC MDC* NNNL **(A)**	**STATA**© *nlogit* NNNL **(A)**	**STATA**© *nlogitrum* RU2 UMNL **(B)**
$\phi_{dc\|A}$	-1.00	-1.18***	-1.18***	-0.72***
$\phi_{cc\|B}$	0.50	0.60***	0.60***	0.24***
$\phi_{dc\|B}$	0.50	0.61***	0.61***	0.24***
$\tau_{dc\|A}$	3.00	3.53***	3.53***	2.13***
$\tau_{cc\|B}$	-1.50	-1.78***	-1.78***	-1.13***
$\tau_{dc\|B}$	-1.00	-1.28***	-1.28***	-0.83***
β_{pro}	2.00	2.41***	2.41***	1.41***
β_{pri}	-1.50	-1.67***	-1.67***	-1.17***
IV_A	0.83	0.51***	0.51***	0.62***
IV_B	0.67	0.63***	0.63***	0.57***
LL		-9,956	-9,956	-9,996

$n = 10,400$; *** $\alpha = 0.01$

5.2 Models With Equality Constraint on the *IV*-Parameters

According to the *utility maximization nested logit* (UMNL) RU2 specification, the scale parameters λ_A and λ_B are set to 1. When simulating data for models C and D (Figure 2), the scale parameters μ_A and μ_B **are** imposed by an equality constraint.
The coefficients of the exogenous variables generated with SAS© are estimated with SAS© and STATA©. Table 9 gives an overview over the simulated data sets for the models C and D. The data sets 2.1 to 2.4 share the

Table 9: Overview data sets for models C and D

	$\mu_A = \mu_B = 1.8$	$\mu_A = \mu_B = 1.6$
input-parameter set #3	**data set 2.1**	**data set 2.3**
input-parameter set #4	**data set 2.2**	**data set 2.4**

fact that the NNNL model estimated with the procedure *PROC MDC* in SAS© results in the same coefficients as the NNNL model estimated with the command *nlogit* in STATA©. The RU2 UMNL model estimated with the command *nlogitrum* in STATA© results in different coefficients. The log-likelihood values and the *IV*-parameters $\left(\frac{\lambda}{\mu}\right)$ of the two *non-normalized nested logit* (NNNL) models and of the RU2 UMNL model are identical.
The *non-normalized nested logit* (NNNL) model estimated with *PROC MDC* in SAS© estimates the same coefficients as the *non-normalized nested logit* (NNNL) model estimated with *nlogit* in STATA©. The RU2 UMNL model estimated with *nlogitrum* in STATA© results in different coefficients. But the coefficients of the two NNNL models and the RU2 UMNL model can be transferred according to Equation (23). The coefficients estimated with STATA© *nlogitrum* equal a multiple of the coefficients estimated with SAS© *PROC MDC* or STATA© *nlogit* respectively. The coefficients estimated in the NNNL models do not have any meaning before their re-scaling, i. e. their multiplication with the estimated *IV*-parameter, and can therefore not be interpreted in the sense of random utility theory. Possible discrepancies

of the coefficients are caused by rounding.
In the data sets 2.1 to 2.4, both SAS© and STATA© estimate the same *IV*-parameter and have the same log-likelihood value. The coefficients estimated with *nlogitrum* in STATA© do not always equal the input-parameters which were used for generating the data. In the data sets 2.1 and 2.3, a reliable reproduction can be recognized. The reasons for a suboptimal parameter reproduction in the data sets 2.2 and 2.4 must be investigated in further research.
The commonly estimated *IV*-parameter 0.77 in data set 2.1 equals a scale parameter of $\mu = \frac{1}{0.77} = 1.30$. Analogously, the scale parameters $\mu = 1.28$ for data set 2.2, $\mu = 1.35$ for data set 2.3, and $\mu = 1.37$ for data set 2.4 result. The scale parameters $\mu = 1.80$ assumed for the simulation of the data sets 2.1 and 2.2, and $\mu = 1.60$ assumed for the simulation of the data sets 2.3 and 2.4, could not be reproduced. But the reproduction of the input-parameters can also be improved by an enlargement of the sample size.
These results show that even with NNNL software a model consistent with random utility theory can be estimated. But the restriction of equal scale parameters and thus equal variances must be accepted.

Table 10: Model estimation with data set 2.1 (RU2 with restriction)

parameter	input-parameter set **#3**	**SAS**© *PROC MDC* NNNL **(C)**	**STATA**© *nlogit* NNNL **(C)**	**STATA**© *nlogitrum* RU2 UMNL **(D)**
$\phi_{dc\|A}$	-0.50	-0.65***	-0.65***	-0.50***
$\phi_{cc\|B}$	0.50	0.60***	0.60***	0.46***
$\phi_{dc\|B}$	0.50	0.60***	0.60***	0.46***
$\tau_{dc\|A}$	1.50	1.87***	1.87***	1.44***
$\tau_{cc\|B}$	-1.00	-1.21***	-1.21***	-0.93***
$\tau_{dc\|B}$	-0.50	-0.63***	-0.63***	-0.49***
β_{pro}	2.00	2.49***	2.49***	1.91***
β_{pri}	-1.50	-1.86***	-1.86***	-1.43***
IV	0.56	0.77***	0.77***	0.77***
LL		-9,724	-9,724	-9,724

$n = 10,400$; *** $\alpha = 0.01$

Table 11: Model estimation with data set 2.2 (RU2 with restriction)

parameter	input-parameter set **#4**	**SAS**© *PROC MDC* NNNL **(C)**	**STATA**© *nlogit* NNNL **(C)**	**STATA**© *nlogitrum* RU2 UMNL **(D)**
$\phi_{dc\|A}$	-1.00	-1.22***	-1.22***	-0.95***
$\phi_{cc\|B}$	0.50	0.56***	0.56***	0.43***
$\phi_{dc\|B}$	0.50	0.57***	0.57***	0.45***
$\tau_{dc\|A}$	3.00	3.62***	3.62***	2.83***
$\tau_{cc\|B}$	-2.00	-2.28***	-2.28***	-1.78***
$\tau_{dc\|B}$	-2.50	-2.95***	-2.94***	-2.30***
β_{pro}	3.00	3.57***	3.57***	2.79***
β_{pri}	-3.00	-3.38***	-3.38***	-2.64***
IV	0.56	0.78***	0.78***	0.78***
LL		-7,636	-7,636	-7,636

$n = 10,400$; *** $\alpha = 0.01$

Table 12: Model estimation with data set 2.3 (RU2 with restriction)

parameter	input-parameter set **#3**	**SAS**© *PROC MDC* NNNL **(C)**	**STATA**© *nlogit* NNNL **(C)**	**STATA**© *nlogitrum* RU2 UMNL **(D)**
$\phi_{dc\|A}$	-0.50	-0.64***	-0.64***	-0.47***
$\phi_{cc\|B}$	0.50	0.58***	0.58***	0.43***
$\phi_{dc\|B}$	0.50	0.58***	0.58***	0.43***
$\tau_{dc\|A}$	1.50	1.85***	1.85***	1.36***
$\tau_{cc\|B}$	-1.00	-1.18***	-1.18***	-0.87***
$\tau_{dc\|B}$	-0.50	-0.61***	-0.61***	-0.45***
β_{pro}	2.00	2.44***	2.44***	1.80***
β_{pri}	-1.50	-1.82***	-1.82***	-1.35***
IV	0.63	0.74***	0.74***	0.74***
LL		-9,940	-9,940	-9,940

$n = 10,400$; *** $\alpha = 0.01$

Table 13: Model estimation with data set 2.4 (RU2 with restriction)

parameter	input-parameter set **#4**	**SAS**© *PROC MDC* NNNL **(C)**	**STATA**© *nlogit* NNNL **(C)**	**STATA**© *nlogitrum* RU2 UMNL **(D)**
$\phi_{dc\|A}$	-1.00	-1.21***	-1.21***	-0.89***
$\phi_{cc\|B}$	0.50	0.55***	0.55***	0.41***
$\phi_{dc\|B}$	0.50	0.57***	0.57***	0.42***
$\tau_{dc\|A}$	3.00	3.58***	3.58***	2.63***
$\tau_{cc\|B}$	-2.00	-2.26***	-2.26***	-1.66***
$\tau_{dc\|B}$	-2.50	-2.92***	-2.92***	-2.14***
β_{pro}	3.00	3.55***	3.55***	2.60***
β_{pri}	-3.00	-3.39***	-3.39***	-2.48***
IV	0.63	0.73***	0.73***	0.73***
LL		-7,840	-7,840	-7,840

$n = 10,400$; *** $\alpha = 0.01$

6 Summary

Although the nested logit model has, because of its ability to account for similarities between alternatives via partial correlation of the error terms, received increasing attention, the various specifications of the nested logit model have barely been focused on. But this differentiation gets its special relevance from the fact that generally only the RU2 UMNL specification is consistent with random utility theory.

Both estimations with real data and simulation studies require investigating the software's underlying nested logit specification. Whereas in estimations with *utility maximization nested logit* (UMNL) software no particularities are to be considered, estimation with *non-normalized nested logit* (NNNL) software proves to be more difficult. Only by imposing restrictions on the *IV*-parameters or by introducing dummy nests can estimation results consistent with random utility theory be reached.

This article also deals with the generation of simulated stochastic utility components. While assumptions on the distribution of the error terms $\nu_{i|m}$ on Level 1 can be directly seen from the extreme-value distribution of the error terms and the nested logit model specification, the assumptions on the distribution of the error terms ν_m on Level 2 requires extensive derivations. This procedure was explained in detail in this article (see 3.4).

It was demonstrated that when using NNNL software without imposing restrictions, a model consistent with random utility theory can **not** be estimated. Three cases are to be distinguished: (1) model without generic coefficients, (2) model with generic coefficients and with equality constraint on the scale parameters, and (3) model with generic coefficients and without equality constraint on the scale parameters. In case (1) the coefficients estimated with NNNL software (e. g. *PROC MDC* in SAS©) can be transferred to the coefficients estimated with UMNL software (e. g. *nlogitrum* in STATA©) by multiplying them with the estimated *IV*-parameter. The thus re-scaled coefficients are the "true" model coefficients. And this is also valid for the coefficients in case (2). A model estimated with NNNL software in case (3) is useless. This becomes especially relevant if the software user is not aware of the described issue of different nested logit model specifications. The danger of a wrong model estimation is then very high.

References

Ailawadi KL, Neslin SA (1998) The Effect of Promotion on Consumption: Buying More and Consuming It Faster. Journal of Marketing Research 35(3):390–398

Anderson SP, de Palma A (1992) Multiproduct Firms: A Nested Logit Approach. The Journal of Industrial Economics 40(3):261–276

Baltas G, Doyle P, Dyson P (1997) A model of consumer choice for national vs private label brands. Journal of the Operational Research Society 48(10):988–995

Ben-Akiva M, Lerman SR (1985) Discrete Choice Analysis: Theory and Application to Travel Demand, sixth printing 1994 edn. The MIT Press, Cambridge/Massachusetts

Bhat CR (1997) Covariance Heterogeneity in Nested Logit Models: Econometric Structure and Application to Intercity Travel. Transportation Research B 31(1):11–21

Cameron AC, Trivedi PK (2005) Microeconometrics. Cambridge University Press

Carrasco JA, de Dios Ortúzar J (2002) Review and assessment of the nested logit model. Transport Reviews 22(2):197–218

Chib S, Seetharaman PB, Strijnev A (2004) Model of Brand Choice with a No-Purchase Option Calibrated to Scanner-Panel Data. Journal of Marketing Research 41(2):184–196

Chintagunta PK (1993) Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households. Marketing Science 12(2):184–208

Chintagunta PK, Vilcassim NJ (1998) Empirical implications of unobserved household heterogeneity for manufacturer and retailer pricing. Journal of Retailing and Consumer Services 5(1):15–24

Daly A (2001) Alternative tree logit models: comments on a paper of Koppelman and Wen. Transportation Research Part B 35(8):717–724

de Dios Ortúzar J (2001) On the development of the nested logit model. Transportation Research Part B 35(2):213–216

Guadagni PM, Little JDC (1983) A Logit Model of Brand Choice Calibrated on Scanner Data. Marketing Science 2(3):203–238

Guadagni PM, Little JDC (1998) When and What to Buy: A Nested Logit Model of Coffee Purchase. Journal of Forecasting 17(3/4):303–326

Heiss F (2002) Structural choice analysis with nested logit models. The Stata Journal 2(3):227–252

Hensher DA, Greene WH (2001) The Mixed Logit Model: The State of Practice and Warnings for the Unwary. Tech. rep., The University of Sydney

Hensher DA, Greene WH (2002) Specification and estimation of the nested logit model: alternative normalisations. Transportation Research Part B 36(1):1–17

Hensher DA, Rose JM, Greene WH (2005) Applied Choice Analysis. Cambridge University Press

Hunt GL (2000) Alternative Nested Logit Model Structures and the Special Case of Partial Degeneracy. Journal of Regional Science 40(1):89–113

Kamakura WA, Kim BD, Lee J (1996) Modeling Preference and Structural Heterogeneity in Consumer Choice. Marketing Science 15(2):152–172

Kannan PK, Wright GP (1991) Modeling and Testing Structured Markets: A Nested Logit Approach. Marketing Science 10(1):58–82

Knapp TA, White NE, Clark DE (2001) A Nested Logit Approach to Household Mobility. Journal of Regional Science 41(1):1–22

Koppelman FS, Wen CH (1998a) Alternative Nested Logit Models: Structure, Properties and Estimation. Transportation Research B 32(5):289–298

Koppelman FS, Wen CH (1998b) Nested Logit Models: Which Are You Using? Transportation Research Record 1645:1–7

McFadden D (1978) Modelling the choice of residential location. In: Karlqvist A, Lundqvist L, Snickars F, Weibull JW (eds) Spatial Interaction Theory and Planning Models, North-Holland Publishing Company, Amsterdam New York Oxford, pp 75–96

McFadden D (1981) Econometric Models of Probabilistic Choice. In: Manski CF, McFadden D (eds) Structural Analysis of Discrete Data with Econometric Applications, The MIT Press, Cambridge, pp 198–272

Munizaga MA, Alvarez-Daziano R (2001) Mixed Logit vs. Nested Logit and Probit Models. Tech. rep., Universidad de Chile

SAS (2004) SAS/ETS 9.1 User's Guide. SAS Institute Inc., Cary/NC, USA

Suárez A, del Bosque IR, Rodríguez-Poo JM, Moral I (2004) Accounting for heterogeneity in shopping centre choice models. Journal of Retailing and Consumer Services 11:119–129

Sun B, Neslin SA, Srinivasan K (2003) Measuring the Impact of Promotions on Brand Switching When Consumers Are Forward Looking. Journal of Marketing Research 40(4):389–405

Train KE (1980) A Structured Logit Model of Auto Ownership and Mode Choice. Review of Economic Studies 47:357–370

Train KE (2003) Discrete Choice Methods with Simulation. Cambridge University Press

SFB 649 Discussion Paper Series 2006

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Calibration Risk for Exotic Options" by Kai Detlefsen and Wolfgang K. Härdle, January 2006.
002 "Calibration Design of Implied Volatility Surfaces" by Kai Detlefsen and Wolfgang K. Härdle, January 2006.
003 "On the Appropriateness of Inappropriate VaR Models" by Wolfgang Härdle, Zdeněk Hlávka and Gerhard Stahl, January 2006.
004 "Regional Labor Markets, Network Externalities and Migration: The Case of German Reunification" by Harald Uhlig, January/February 2006.
005 "British Interest Rate Convergence between the US and Europe: A Recursive Cointegration Analysis" by Enzo Weber, January 2006.
006 "A Combined Approach for Segment-Specific Analysis of Market Basket Data" by Yasemin Boztuğ and Thomas Reutterer, January 2006.
007 "Robust utility maximization in a stochastic factor model" by Daniel Hernández–Hernández and Alexander Schied, January 2006.
008 "Economic Growth of Agglomerations and Geographic Concentration of Industries - Evidence for Germany" by Kurt Geppert, Martin Gornig and Axel Werwatz, January 2006.
009 "Institutions, Bargaining Power and Labor Shares" by Benjamin Bental and Dominique Demougin, January 2006.
010 "Common Functional Principal Components" by Michal Benko, Wolfgang Härdle and Alois Kneip, Jauary 2006.
011 "VAR Modeling for Dynamic Semiparametric Factors of Volatility Strings" by Ralf Brüggemann, Wolfgang Härdle, Julius Mungo and Carsten Trenkler, February 2006.
012 "Bootstrapping Systems Cointegration Tests with a Prior Adjustment for Deterministic Terms" by Carsten Trenkler, February 2006.
013 "Penalties and Optimality in Financial Contracts: Taking Stock" by Michel A. Robe, Eva-Maria Steiger and Pierre-Armand Michel, February 2006.
014 "Core Labour Standards and FDI: Friends or Foes? The Case of Child Labour" by Sebastian Braun, February 2006.
015 "Graphical Data Representation in Bankruptcy Analysis" by Wolfgang Härdle, Rouslan Moro and Dorothea Schäfer, February 2006.
016 "Fiscal Policy Effects in the European Union" by Andreas Thams, February 2006.
017 "Estimation with the Nested Logit Model: Specifications and Software Particularities" by Nadja Silberhorn, Yasemin Boztuğ and Lutz Hildebrandt.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".