Discrete-continuous maximum likelihood for the estimation of nested logit models

Anna Fernández Antolín

Virginie Lurkin

Matthieu de Lapparent

Michel Bierlaire

Transportation and Mobility Laboratory, ENAC, EPFL

May 2017

Transportation and Mobility Laboratory, ENAC, EPFL

Discrete-continuous maximum likelihood for the estimation of nested logit models

Anna Fernández Antolín, Virginie Lurkin, Matthieu de Lapparent, Michel Bierlaire
Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne
Station 18
CH-1015 Lausanne
phone: +41-21-693 24 35
fax: +41-21-693 80 60
{anna.fernandezantolin,virginie.lurkin,matthieu.delapparent,michel.bierlaire}@epfl.ch

May 2017

Abstract

In this paper we aim at integrating the selection of a nesting structure to the maximum likelihood framework of the parameter estimation. Given a finite set of nesting structures, the traditional approach is to estimate the models corresponding to each of them and select a posteriori the most appropriate one based on some fit statistics and informal testing procedures. However, the number of possible nesting structures grows as a function of the number of alternatives.

Our approach simultaneously solves the problem of selecting the optimal nesting structure and estimating its corresponding parameters with maximum likelihood. We call this *discrete-continuous maximum likelihood* (DCML). We are able to linearize the logarithm in the objective function so that it results in a mixed integer linear problem.

Keywords

discrete-continuous maximum likelihood, nested logit model, mixed integer linear program

1 Introduction

In the family of discrete choice models, the multinomial logit (MNL) model is the most frequently used in practice. Its key advantage relies on the simplicity of its closed-form probability expression. In the MNL model, the error terms are supposed to be independently and identically distributed across alternatives and individuals. This leads to the independence of irrelevant alternatives (IIA) property, a property which states that the ratio of two choice probabilities is independent of the attributes or even the existence of any other alternatives. In terms of substitution patterns, the IIA property implies unrealistic substitution patterns across alternatives.

To overcome this limitation, more advanced models, that incorporate more realistic substitution patterns by relaxing the assumptions of independent errors terms, have been proposed. Among these more advanced models, the nested, cross nested and mixed logit models have been extensively studied in the literature, mostly from a theoretical point of view. The common idea behind these models is to allow patterns of correlation by placing the alternatives into several groups called nests. Alternatives that belong to the same nest share a common error term and are therefore correlated, while alternatives that are not in the same nest are independent. We refer the interested reader to Bierlaire (2006) and Abbe *et al.* (2007) for a detailed analysis of the cross-nested logit model (CNL) and to Walker *et al.* (2007) for an analysis on error component logit-mixture models.

In some cases, it is clear which alternatives share unobserved attributes and the nesting structure is obvious. However, in other cases there are several nesting structures that make intuitive sense. In practice, to determine the most appropriate nesting structure, the analyst has several options. The first option is to estimate a model for each possible nesting structure and to choose a posteriori the one which best fits the data. Another option is to use a cross nested structure where all alternatives belong to all nests and the membership parameters of each alternative to each nest are estimated. However, the parameters of cross-nested logit models are often difficult to estimate due to the non-convexity of the likelihood function.

We therefore propose to find the best –in terms of likelihood– nested logit model by introducing binary variables that define the allocation of alternatives to nests. We fix the number of nests and find the best nesting structure by estimating simultaneously the binary variables and the continuous parameters in the utility functions. We build on the framework developed by Pacheco *et al.* (forthcoming) and show that we can formulate the problem defined above as a mixed integer linear problem (MILP). To the best of our knowledge, this is the first time that discrete variables are used in the context of discrete choice models within the maximum likelihood framework. The contributions of the paper are both introducing *discrete-continuous maximum likelihood* (DCML) and linearizing the log-likelihood function by relying on simulation.

The remaining of the paper is organized as follows. The mathematical model is presented in Section 2, followed by the case study in Section 3. Finally, the conclusions of the paper and future research directions are presented in Section 4.

2 Mathematical model

In order to model the nesting structure, we use a nested logit model with the following utility functions

$$U_{in} = V_{in} + \varepsilon_{in}, \tag{1}$$

where V_{in} is the deterministic part of the utility function for alternative i and individual n. V_{in} is a linear-in-parameters function of a vector of parameters to be estimated (β), observed attributes of the alternatives (a_{in}) and socioeconomic characteristics of the individual (s_n), $V_{in} = f(a_{in}, s_n, \beta)$. ε_{in} are the error terms, that capture the correlation between alternatives. The choice set is denoted as C.

Following Ben-Akiva and Lerman (1985), the error term associated with each alternative i that belongs to nest m can be decomposed into a common error component, ε_{mn}, and an independent error term, ε_{imn}

$$\varepsilon_{in} = \varepsilon_{mn} + \varepsilon_{imn}, \tag{2}$$

where

- ε_{mn} is such that $\tilde{\varepsilon}_{mn} = \varepsilon_{mn} + \varepsilon'_{mn}$, where $\tilde{\varepsilon}_{mn} \overset{iid}{\sim} EV(0, \mu)$ and $\varepsilon'_{mn} \overset{iid}{\sim} EV(0, \mu_m)$.
- $\varepsilon_{imn} \overset{iid}{\sim} EV(0, \mu_m)$,

Therefore, Equation (1) can be rewritten as

$$U_{in} = V_{in} + \tilde{\varepsilon}_{mn} + (\varepsilon_{imn} - \varepsilon'_{mn}), \tag{3}$$

where $\tilde{\varepsilon}_{mn} \overset{iid}{\sim} EV(0, \mu)$, $\varepsilon'_{mn} \overset{iid}{\sim} EV(0, \mu_m)$, $\varepsilon_{imn} \overset{iid}{\sim} EV(0, \mu_m)$. For normalization reasons, $\mu = 1$ and $\mu_m \geq \mu = 1$.

From the properties of the extreme value distribution, we know that if $\varepsilon_{imn} \sim EV(0, \mu_m)$, then

$\xi_{imn} = \frac{1}{\mu_m}\varepsilon_{imn} \sim EV(0,1)$. Analogously, if $\varepsilon'_{mn} \sim EV(0,\mu_m)$, then $\xi'_{mn} = \frac{1}{\mu_m}\varepsilon'_{mn} \sim EV(0,1)$.

and Equation (3) can be rewritten as

$$U_{in} = V_{in} + \tilde{\varepsilon}_{mn} + \frac{1}{\mu_m}(\xi_{imn} - \xi'_{mn}). \tag{4}$$

Finally, as we don't know a priori if alternative i belongs to nest m, we introduce the following binary decision variables:

$$b_{im} = \begin{cases} 1 & \text{if alternative } i \text{ belongs to nest } m, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

and our utility function (4) becomes:

$$U_{in} = V_{in} + \sum_{m=1}^{M} b_{im}\left(\tilde{\varepsilon}_{mn} + \frac{1}{\mu_m}(\xi_{imn} - \xi'_{mn})\right) \tag{6}$$

$$= V_{in} + \sum_{m=1}^{M} b_{im}\tilde{\varepsilon}_{mn} + \sum_{m=1}^{M}\left(\frac{b_{im}}{\mu_m}(\xi_{imn} - \xi'_{mn})\right). \tag{7}$$

2.1 Objective function

In order to apply maximum likelihood, we want to maximize the following function

$$\log\left(\prod_{n=1}^{N}\prod_{i=1}^{I} P_n(i)^{d_{in}}\right), \tag{8}$$

where $P_n(i)$ is the probability that individual n chooses alternative i, and d_{in} is an observed variable that takes value 1 if individual n chooses alternative i and 0 otherwise. N is the number of individuals and I is the number of alternatives, $C = \{1, 2, \ldots, I\}$. The probability that individual n chooses alternative i within her choice set is

$$P_n(i) = P(U_{in} \geq U_{jn},\ \forall j \neq i). \tag{9}$$

Using the framework developed by Pacheco *et al.* (forthcoming) we can use simulation of the error terms and avoid the non-linearities caused by the expression of the probabilities. This is done by working with the values of the utility functions instead of with the probabilities. We generate, for each error term, R draws based on the distributional assumptions. Once the draws have been generated, we obtain the utility associated with each alternative i by each individual n

in each scenario $r = 1, \dots, R$

$$U_{inr} = V_{in} + \sum_{m=1}^{M} b_{im}\tilde{\varepsilon}_{mnr} + \sum_{m=1}^{M}\left(\frac{b_{im}}{\mu_m}(\xi_{imnr} - \xi'_{mnr})\right). \tag{10}$$

Then the objective function becomes

$$\sum_{n=1}^{N}\sum_{i=1}^{I} d_{in}\log\left(\frac{1}{R}\sum_{r=1}^{R} w_{inr}\right), \tag{11}$$

where

$$w_{inr} = \begin{cases} 1 & \text{if} \quad U_{inr} > U_{jnr} \quad \forall j \neq i, \\ 0 & \text{otherwise,} \end{cases} \qquad \forall i, n, r. \tag{12}$$

The only remaining non-linearity is the logarithm that appears in the objective function. Since $\sum_{r=1}^{R} w_{inr}$ can only take integer values from 1 to R, we can linearize it by introducing binary variables denoted γ_{inp} defined as follows

$$\gamma_{inp} = \begin{cases} 1 & \text{if } \sum_{r=1}^{R} w_{inr} = p, \\ 0 & \text{otherwise,} \end{cases} \qquad \forall i, n, p. \tag{13}$$

Then, Equation (11) is equivalent to

$$\sum_{n=1}^{N}\sum_{i=1}^{I} d_{in}\sum_{p=1}^{R}\gamma_{inp}L_p, \tag{14}$$

where $L_p = \log(p),\ p = 1, ..., R$ is a pre-processed vector of R components. The linearization of Equation (13) is described in Section 2.2.

2.2 Constraints

Linearization of the utility functions In order to linearize Equation (10), we define variables $\bar{\mu}_m = \frac{1}{\mu_m} \in (0, 1]$ and $\tau_{im} = b_{im}\bar{\mu}_m$. Then the linearization of Equation (10) is as follows

$$U_{inr} = f(a_{in}, s_n, \beta) + \sum_{m=1}^{M} b_{im}\tilde{\varepsilon}_{mnr} + \sum_{m=1}^{M} \tau_{im}(\xi_{imnr} - \xi'_{mnr}), \qquad \forall i, n, r, \tag{15}$$

$$\tau_{im} \leq b_{im}, \qquad \forall i, m, \tag{16}$$

$$\tau_{im} \leq \bar{\mu}_m, \qquad \forall i, m, \tag{17}$$

$$\tau_{im} \geq \bar{\mu}_m + b_{im} - 1, \qquad \forall i, m. \tag{18}$$

Discounted utility We introduce the variable that denotes the availability of an alternative i for an individual n as follows

$$y_{in} = \begin{cases} 1 & \text{if alternative } i \text{ is available for individual } n, \\ 0 & \text{otherwise,} \end{cases} \qquad \forall i, n. \tag{19}$$

The discounted utility associated to individual n, alternative i and draw r is denoted z_{inr} and defined as

$$z_{inr} = \begin{cases} U_{inr} & \text{if } y_{in} = 1, \\ \ell_{nr} & \text{if } y_{in} = 0, \end{cases} \qquad \forall i, n, r. \tag{20}$$

where $\ell_{nr} = \min_{j \in C} \ell_{jnr}$ is the smallest lower bound across all alternatives. The linear formulation of Equation (20) is

$$l_{nr} \leq z_{inr}, \qquad \forall i, n, r, \tag{21}$$

$$z_{inr} \leq l_{nr} + M_{inr}y_{in}, \qquad \forall i, n, r, \tag{22}$$

$$U_{inr} - M_{inr}(1 - y_{in}) \leq z_{inr}, \qquad \forall i, n, r, \tag{23}$$

$$z_{inr} \leq U_{inr}, \qquad \forall i, n, r, \tag{24}$$

as shown by Pacheco *et al.* (forthcoming).

Choice As defined in Equation (12), w_{inr} is a binary variable that takes value 1 if alternative i is chosen by individual n in draw r and zero otherwise. The following constraints model the fact that an individual can only choose one alternative per draw and that only available alternatives can be chosen by an individual

$$\sum_{i=1}^{I} w_{inr} = 1, \qquad \forall n, r, \tag{25}$$

$$w_{inr} \leq y_{in}, \qquad \forall i, n, r. \tag{26}$$

Moreover, based on the behavioral assumption, the chosen alternative of an individual corresponds to its associated highest discounted utility. We introduce the continuous variable U_{nr} that is defined as

$$U_{nr} = \max_{i \in C} z_{inr}, \quad \forall n, r. \tag{27}$$

Its linear formulation (Pacheco *et al.* (forthcoming)) is given by

$$z_{inr} \leq U_{nr}, \qquad \forall i, n, r, \tag{28}$$

$$U_{nr} \leq z_{inr} + M_{nr}(1 - w_{inr}) \qquad \forall i, n, r, \tag{29}$$

where $m_{nr} = \max_{j \in C} m_{jnr}$ is the largest upper bound accross all alternatives, and $M_{nr} = m_{nr} - \ell_{nr}$ is the difference between the largest upper bound and the smallest lower bound.

Linearization of γ_{inp} Equation (13) can be linearized as follows

$$(R+1)\delta^1_{inp} - 1 \geq \sum_{r=1}^{R} w_{inr} - p, \qquad \forall i, n, p, \tag{30}$$

$$(R+1)\delta^2_{inp} - 1 \geq p - \sum_{r=1}^{R} w_{inr}, \qquad \forall i, n, p, \tag{31}$$

$$\delta^1_{inp} + \delta^2_{inp} - 2\gamma_{inp} \leq 1, \qquad \forall i, n, p, \tag{32}$$

$$\sum_{p=1}^{R} \gamma_{inp} = 1, \qquad \forall i, n, \tag{33}$$

where $\delta^1_{inp}, \delta^2_{inp}$ are binary variables. To prove the equivalence between Equation (13) and Equations (30)-(33) we consider three cases:

- If $\sum_{r=1}^{R} w_{inr} = p$, constraints (30)-(31) become

$$(R+1)\delta^1_{inp} - 1 \geq 0 \qquad \forall i, n, p, \tag{34}$$

$$(R+1)\delta^2_{inp} - 1 \geq 0 \qquad \forall i, n, p. \tag{35}$$

Constraints (34) and (35) impose that $\delta^1_{inp} = \delta^2_{inp} = 1$. Using this, constraint (32) is written

$$2 - 2\gamma_{inp} \leq 1 \iff 1 \leq 2\gamma_{inp} \iff \gamma_{inp} = 1. \tag{36}$$

From constraint (33), $\gamma_{inr} = 0$ if $r \neq p$.

- If $\sum_{r=1}^{R} w_{inr} > p$, constraint (30) becomes

$$(R+1)\delta^1_{inp} - 1 \geq \sum_{r=1}^{R} w_{inr} - p > 0 \iff (R+1)\delta^1_{inp} > 1 \iff \delta^1_{inp} = 1 \tag{37}$$

From constraint (33) we obtain that $\gamma_{inp} = 0$, so from constraint (32) $\delta^2_{inp} = 0$ and constraint (32) is trivial.

- If $\sum_{r=1}^{R} w_{inr} < p$, the derivation is analogous to the previous case.

Nesting structure To express that each alternative belongs to exactly one nest we use the following constraint

$$\sum_{m=1}^{M} b_{im} = 1, \quad \forall i \tag{38}$$

and in order to break possible symmetries

$$b_{im} = 0, \quad \forall m > i. \tag{39}$$

Finally, for identification purposes, the scale of a nest with only one alternative must be 1. That is, if $\sum_{i=1}^{I} b_{im} = 1$ then $\bar{\mu}_m = 1$, $\forall m$. This implication is linearized by the following constraints

$$\sum_{i=1}^{I} b_{im} \leq Mq_m, \qquad \forall m, \tag{40}$$

$$2 - \sum_{i=1}^{I} b_{im} \leq Mt_m, \qquad \forall m, \tag{41}$$

$$q_m + t_m \leq 1 + \bar{\mu}_m, \qquad \forall m, \tag{42}$$

where t_m and q_m are binary variables and M a big number. To prove the equivalence we consider the following:

- If $\sum_{i=1}^{I} b_{im} = 1$, constraints (40)-(41) become

$$1 \leq Mq_m, \qquad \forall m, \tag{43}$$

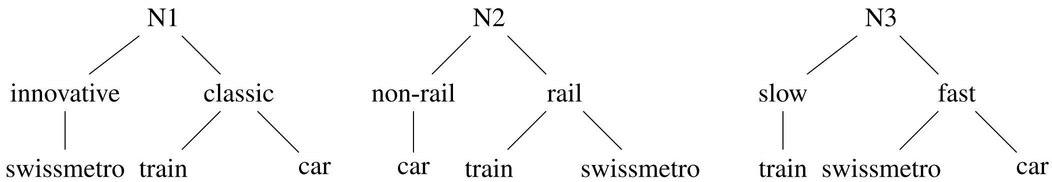
$$1 \leq Mt_m, \qquad \forall m,. \tag{44}$$

From constraint (43) we have that $q_m = 1$, and from constraint (44) we have that $t_m = 1$. Then, constraint (42) becomes $2 \leq 1 + \bar{\mu}_m \iff 1 \leq \bar{\mu}_m$. Since by definition, $\bar{\mu}_m \in (0, 1]$, we obtain that $\bar{\mu}_m = 1$.

3 Case study

For the proof-of-concept we use a stated preferences mode choice case study collected in Switzerland in 1998. The respondents provided information in order to analyze the impact of the model innovation in transportation represented by the Swissmetro, a mag-lev underground system, compared to the usual transport modes of car and train.

The choice set of the respondents is C = {car, train, swissmetro}. A possible assumption is that the modes *train* and *car* share unobserved attributes due to the fact that they are both *classic* or *existing* transportation modes. This is represented by Figure 1(a). One could also hypothesize what is represented in Figure 1(b), that it is the alternatives *train* and *swissmetro* that share unobserved attributes, due to the fact that they are both rail-based, unlike the *car* alternative. Finally, it could also be that *swissmetro* and *car* are correlated due to the fact that they are generally faster than the *train* alternative.



((a)) Innovative vs. classic ((b)) Rail vs. non-rail ((c)) Fast vs. slow

Figure 1: Possible nesting structures with two nests.

Table 1 shows the model specification considered.

Parameter	Car	Train	Swissmetro
ASC_{CAR}	1	0	0
ASC_{SM}	0	0	1
β_{COST}	COST_{car}	$\text{COST}_{\text{train}}$	COST_{sm}
β_{HE}	0	HE_{train}	HE_{sm}
β_{TIME}	TIME_{car}	$\text{TIME}_{\text{train}}$	TRAIN_{sm}

Table 1: Model specification - Deterministic part of the utility functions

Enumeration of the nesting structures In order to have a benchmark we first estimate the models discussed above using pythonbiogeme (Bierlaire (2016)). The results, with the nesting structures introduced in Figure 1, are presented in Table 2. Models N2 and N3 are not identified, since both μ_{RAIL} and μ_{FAST} reach the lower bounds. For this reason, the rest of the parameters, as well as the final log likelihood is the same for both models. From these results we conclude that the best model is N1. It is also better than an multinomial logit model (MNL) since $\mu_{CLASSIC}$ is significantly different from one.

	N1		N2		N3	
Parameter	Value	t-test	Value	t-test	Value	t-test
ASC_{CAR}	0.0943	1.71	0.189	2.37	0.189	2.37
ASC_{SM}	0.335	4.04	0.451	4.84	0.451	4.84
β_{COST}	-0.00860	-14.38	-0.0108	-15.90	-0.0108	-15.90
β_{HE}	-0.00380	-5.45	-0.00535	-5.45	-0.00535	-5.45
β_{TIME}	-0.00900	-8.38	-0.0128	-12.23	-0.0128	-12.23
$\mu_{CLASSIC}$	2.06	6.50 [1]	-	-	-	-
μ_{RAIL}	-	-	1.00	0.00[1]	-	-
μ_{FAST}	-	-	-	-	1.00	0.00[1]
$\mathcal{L}(\hat{\beta})$	-5219.883		-5315.386		-5315.386	

Table 2: Estimation results: biogeme

The next step is to obtain results from the framework defined in Section 2, and compare them to the results from Table 2.

4 Conclusion

We have introduced the concept of *discrete-continuous maximum likelihood* and shown that it can be modeled as a mixed integer linear program. This framework allows to simultaneously estimate the (continuous) parameters of the utility function as well as the (discrete) allocation parameters of alternatives to nests.

Moreover, we have found a linear approximation of the log-likelihood function. This framework can therefore also be used to insure a global optimum in other discrete choice models where

[1] t-tests for *mu* parameters are against one.

exact expression of the log-likelihood is non-convex, such as cross-nested logit models, when the nesting structure of the alternatives is known.

We consider a case study with three alternatives so that a full enumeration of the nesting structures is possible, in order to have a benchmark for the results that we will obtain. The data used is a stated preference mode choice case study developed in Switzerland in 1998. This research is still in progress. The next step is to obtain results using the MILP framework developed in order to compare them with the results from the full enumeration of nesting structures and to show that our approach is numerically feasible.

5 References

Abbe, E., M. Bierlaire, and T. Toledo (2007) Normalization and correlation of cross-nested logit models, *Transportation Research Part B: Methodological*, **41** (7) 795–808.

Ben-Akiva, M. and S. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press series in transportation studies, MIT Press, ISBN 978-0-262-02217-0.

Bierlaire, M. (2006) A theoretical analysis of the cross-nested logit model, *Annals of Operations Research*, **144** (1) 287–300.

Bierlaire, M. (2016) Pythonbiogeme: a short introduction, *Technical Report*, **TRANSP-OR 160706, Series on Biogeme**, Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Pacheco, M., S. S. Azadeh, M. Bierlaire and B. Gendron (forthcoming) Integrating advanced discrete choice models in mixed integer linear programming, *Technical Report*, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

Walker, J. L., M. Ben-Akiva and D. Bolduc (2007) Identification of parameters in normal error component logit-mixture (NECLM) models, *Journal of Applied Econometrics*, **22** (6) 1095–1125, September 2007, ISSN 08837252, 10991255.

A Complete model

The discrete continuous maximum likelihood problem can be formalized as follows:

$$
\begin{aligned}
\max \quad & \sum_{n=1}^{N}\sum_{i=1}^{I} d_{in} \sum_{p=1}^{R} \gamma_{inp} L_p & & \\
\text{subject to} \quad & U_{inr} = f(a_{in}, s_{in}, \beta) + \sum_{m=1}^{M} b_{im}\tilde{\varepsilon_{mnr}} + \sum_{m=1}^{M} \tau_{im}(\xi_{imnr} + \xi'_{mnr}) & \forall i,n,r \quad & (45) \\
& w_{inr} \le y_{in} & \forall i,n,r \quad & (46) \\
& l_{nr} \le z_{inr} & \forall i,n,r \quad & (47) \\
& z_{inr} \le l_{nr} + M_{inr} y_{in} & \forall i,n,r \quad & (48) \\
& U_{inr} - M_{inr}(1 - y_{in}) \le z_{inr} & \forall i,n,r \quad & (49) \\
& z_{inr} \le U_{inr} & \forall i,n,r \quad & (50) \\
& z_{inr} \le U_{nr} & \forall i,n,r \quad & (51) \\
& U_{nr} \le z_{inr} + M_{nr}(1 - w_{inr}) & \forall i,n,r \quad & (52) \\
& \sum_{i=1}^{I} w_{inr} = 1 & \forall n,r \quad & (53) \\
& \sum_{m=1}^{M} b_{im} = 1 & \forall i \quad & (54) \\
& b_{im} = 0 & \forall m > i \quad & (55) \\
& \tau_{im} \le b_{im} & \forall i,m \quad & (56) \\
& \tau_{im} \le \bar{\mu}_m & \forall i,m \quad & (57) \\
& \tau_{im} \ge \bar{\mu}_m + b_{im} - 1 & \forall i,m \quad & (58) \\
& \sum_{i=1}^{I} b_{im} \le M s_m & \forall m \quad & (59) \\
& 2 - \sum_{i=1}^{I} b_{im} \le M t_m & \forall m \quad & (60) \\
& s_m + t_m \le 1 + \bar{\mu}_m & \forall m \quad & (61) \\
& (R+1)\delta^1_{inp} - 1 \ge \sum_{r=1}^{R} w_{inr} - p & \forall i,n,p \quad & (62) \\
& (R+1)\delta^2_{inp} - 1 \ge p - \sum_{r=1}^{R} w_{inr} & \forall i,n,p \quad & (63) \\
& \delta^1_{inp} + \delta^2_{inp} - 2\gamma_{inp} \le 1 & \forall i,n,p \quad & (64) \\
& \sum_{p=1}^{R} \gamma_{inp} = 1 & \forall i,n \quad & (65) \\
& s_m, t_m \in \{0,1\} & \forall m \quad & (66) \\
& b_{im} \in \{0,1\} & \forall i,m \quad & (67) \\
& w_{inr}, \delta^1_{inr}, \delta^2_{inr} \in \{0,1\} & \forall i,n,r \quad & (68) \\
& \gamma_{inp} \in \{0,1\} & \forall i,n,p \quad & (69) \\
& \bar{\mu}_m \in [0,1] & \forall m \quad & (70) \\
& \tau_{im} \in [0,1] & \forall i,m \quad & (71) \\
& \beta \in R^p & & (72)
\end{aligned}
$$

- (45) are the utility functions.
- (46)-(48) are used to linearize $b_{im}\sigma_m$.
- (49)-(52) are equivalent to $z_{inr} = U_{inr}$ if $y_{in} = 1$, $z_{inr} = l_{nr}$ if $y_{in} = 0$, which sets the utility of a given alternative to a lower bound if the alternative is not available.
- (53) express the fact that each customer chooses one alternative.
- (54) say that only available alternatives can be selected by individuals.
- (55)-(56) are equivalent to $U_{nr} = \max_i z_{inr}$.
- Constraints (57)-(60) are equivalent to $\gamma_{inp} = 1 \iff \sum_{r=1}^{R} w_{inr} = p$.
- (61) say that each alternative belongs to exactly one nest.
- (62) are used as a symmetry-breaking constraints.
- (63) to (68) define the space of solutions.