Local Relaxation of Constraints on Dissimilarity Parameters in the Generalized Nested Logit Model
Kei TAKAHASHI*a

*aCenter for Mathematics and Data Science, Gunma University, Gunma, 371-8510 Japan

Abstract
This paper derives the exact necessary and sufficient constraints on the dissimilarity coefficients in the Generalized Nested Logit (GNL) model. Previous studies have assumed the Nested Logit (NL) model’s dissimilarity coefficients lie within the unit interval, consistent with random utility maximization. However, those previous studies relaxed constraints on dissimilarity coefficients only for the NL model. We derived constraints on the dissimilarity coefficients for the GNL model. We tested the consistency by deriving the constraints of the NL model from those of the GNL model. Then, we proposed the parameter estimation method under derived constraints. We performed the parameter estimation using POS data from supermarkets. As a result, we obtained higher likelihoods and catching rates compared with the previous conditions.

Key words: discrete choice model, the generalized nested logit mode, marketing, demand forecasting

1 Introduction
Almost all actual purchasing behaviors are under discrete choice. Huge studies have been conducted to capture discrete purchasing behavior through discrete choice models. Some discrete choice models are part of the Generalized Extreme Value (GEV) family of models, which are consistent with random utility maximization (RUM). The GEV models include the Nested Logit (NL) model that has a tree structure to treat similarities between alternatives, the Cross Nested Logit and Generalized Nested Logit (GNL) models[1] that expand the NL model to treat complex correlations among alternatives. In GEV models, conditions on parameters called Daly-Zachary-McFadden (DZM) conditions[2, 3] are needed for consistency with RUM. Many studies conducted parameter estimations under DZM conditions, which are sufficient, but not necessary and sufficient. Then, if DZM conditions are relaxed, it is possible to obtain estimates that are more preferable.

This paper exhibits constraints relaxing DZM conditions in the GNL model, which are necessary and sufficient. Then, parameter estimations are conducted under these constraints. For concrete, local conditions on dissimilarity coefficients in the GNL model are derived analytically, where “local” means conditional on concrete choice structure. Then, we propose methods for parameter estimation under these constraints and compare them to ones under DZM conditions with actual ID-POS data. The GNL model includes the NL and CNL models; therefore, relaxing constraints in the GNL model has profound significance in understanding the whole GEV family. Some relevant studies[4, 5, 6, 7] tackle the relaxation of DZM conditions in the NL model, but no studies have tackled the relaxation of DZM conditions in the GNL model. In GNL models, it is possible for alternatives to belong to plural nests. Then, it is easy to understand that relaxed constraints in the GNL model are complex, and different from the ones in the NL model.

This paper is organized as follows. Section 2 discusses GEV models and random utility maximization, with the GNL model as the preliminary. Section 3 includes the main results, deriving the relaxed constraints, and proposing the parameter estimation method under those constraints. Section 4 applies the proposed methods to estimate parameters for actual data. Finally, Section 5 presents our conclusions.

2 Preliminary

2.1 The GNL Model
The GNL model has a hierarchical structure. In the GNL model, when consumer i chooses alternative n, the choice...
probability \( Q^i_n \) is represented as
\[
Q^i_n = \sum_{m=1}^{N_n} P^i_m P^i_{n|m}, \tag{1}
\]
\[
P^i_m = \sum_{m=1}^{N_m} \left( \sum_{k \in K_m} (\alpha_{km} \exp V^i_k) \right) \mu_m,
\]
\[
P^i_{n|m} = \frac{(\alpha_{nm} \exp V^i_n) \mu_m}{\sum_{k \in K_m} (\alpha_{km} \exp V^i_k)} \mu_m,
\tag{3}
\]
where \( P^i_m \) is the probability of consumer \( i \) choosing nest \( m \), and \( P^i_{n|m} \) is the conditional probability of consumer \( i \) choosing alternative \( n \) after having chosen nest \( m \). \( V^i_n \) denotes the (definite) utility of alternative \( n \) for consumer \( i \). \( N_n \) is the total number of nests in the same GNL model, and \( K_m \) is the alternative set belonging to nest \( m \). The GNL model is parameterized by two types of parameters: dissimilarity and allocation coefficients. \( \mu_m \) denotes the dissimilarity coefficient for nest \( m \) and \( \alpha_{nm} \) denotes the allocation coefficient for alternative \( n \) in nest \( m \). The subscript \( i \) stands for consumer \( i \) omitted below if it is not necessary, because it is confused with exponents.

For global consistency with RUM, \( \mu_m \) and \( \alpha_{nm} \) need to satisfy the following conditions:
\[
0 < \mu_m \leq 1 \quad \forall m, \tag{4}
\]
\[
\alpha_{nm} \geq 0 \quad \forall (n, m), \tag{5}
\]
\[
\sum_{n=1}^{N_n} \alpha_{nm} = 1 \quad \forall n. \tag{6}
\]
As \( \mu_m \) approaches zero, choice probabilities in the same nest have higher correlations, and vice versa as \( \mu_m \) approaches one. This fact is also interpreted by the quasi-correlation coefficient shown in Papola[8]:
\[
\rho_{nn'} \approx \sum_{n=1}^{N_n} \alpha_{nn'}^{1/2} \alpha_{nn'}^{1/2} (1 - \mu_m^2), \quad \forall n, n'. \tag{7}
\]
where \( \rho_{nn'} \) is the correlation coefficient between choice probabilities for alternatives \( n \) and \( n' \). Equation (7) shows that their correlation coefficient is approximated by \( 1 - \mu_m^2 \), with the weight \( \alpha_{nn'} \alpha_{nn'}^{1/2} \) when alternatives \( n \) and \( n' \) belong to the same nest \( m \).

2.2 RUM in the GNL Model

Choice probability in many discrete choice models can be derived from an assumption of RUM. RUM has strict consistency with utility maximization via aggregation of consumer surplus. In the case of the GNL model, we can comprehend consumer behavior as the following minimax problem. The choice probabilities shown in equations (1)–(6) are equivalent to the solution of the stochastic maximization problem with entropy constraints[12]:
\[
E = \max_{P^i_m|n} \left[ \sum_{k=1}^{N_n} \sum_{m=1}^{N_m} P^i_n P^i_{n|m} \ln \frac{P^i_{n|m}}{\alpha_{km}} \right] \tag{8}
\]
where \( s.t. \) \( \sum_{m=1}^{N_m} \sum_{k \in K_m} P^i_m = 1, \)
\[
\text{if } P^i_m = 0 \text{ or } \alpha_{km} = 0 \quad \implies P^i_{n|m} \ln \frac{P^i_{n|m}}{\alpha_{km}} = 0, \quad \forall k, m.
\]
The first term of equation (8) indicates RUM behavior. The second and third terms are entropy constraints in the nest and the alternative levels, respectively, and their ratio of second terms to third terms is approximately \( \mu_m : 1 - \mu_m \). The other conditions are for \( P^i_{n|m} \) as probability and specific dealing at zero. Therefore, the GNL model shown in equations (1)–(3) is consistent with RUM.

2.3 DZM Conditions and the GEV model

The Daly-Zachary-McFadden conditions, derived by Daly and Zachary[2] and McFadden (1981)[3], are the ones discrete choice models must meet to be consistent with RUM. The NL and GNL models, when all dissimilarity coefficients lie in the unit interval, are global consistent with RUM. However, Börsch-Supan[4, 5] refers to the DZM conditions as global and sufficient conditions. Therefore, if one specifies choice structure, the DZM conditions can be relaxed to the local, necessary, and sufficient conditions. One Family that satisfies the DZM conditions is the Generalized Extreme Value (GEV).

The GEV Family or Model is advocated by McFadden (1978)[9]. Each GEV model has a corresponding GEV generating function. When the GEV generating function satisfies the following four conditions, it is consistent with RUM.

We define \( Y_n = \exp V_n \) and set the number of alternatives \( N \). Then, the GEV generating function takes the form \( G = G(Y_1, \ldots, Y_N) \). After the partial differential of \( G \) by \( Y_n \), we obtain the choice probability of alternative
Local Relaxation of Constraints on Dissimilarity Parameters in the Generalized Nested Logit Model

\[ Q_n = \frac{Y_n \partial G}{G \partial Y_n}. \]  

If \( G \) satisfies the following four conditions, then the choice probability \( Q_n \) corresponding to \( G \) is global consistent with RUM.

**Condition 1.** \( G(Y_1, \ldots, Y_N) \geq 0. \)

**Condition 2.** \( G \) is a \( q \)-degree homogeneous function, and for any arbitrary real number \( r \), \( G(rY_1, \ldots, rY_N) = r^q G(Y_1, \ldots, Y_N). \)

**Condition 3.** Mutual partial derivatives of \( G \) with respect to an arbitrary pair of \( k \) in \( Y_n \) become non-negative when \( k \) is an odd number, otherwise they become non-positive.

When the appropriate GEV generating function satisfies conditions 1–4, then the GNL model shown in (1)–(3) with global constraints (4)–(6) can be derived.

### 2.4 Relaxation in the NL model

Herriges and Kling\([7]\) derived the upper limits of dissimilarity coefficients in relaxed constraints shown in Börsch-Supan, which were corrected by Herriges and Kling. They showed the following first- and second-order conditions:

\[ \tau_m \leq P_m, \]
\[ 2(\tau_m - P_m)^2 + \tau_m P_m \geq \tau_m. \]

where, \( \tau_m \equiv \frac{\mu_m - 1}{P_m}. \)

Kling and Herriges\([6]\) conducted empirical analysis with actual data under Herriges and Kling conditions. They compared statistics under DZM conditions to that of choice data of fishing places. They limited the application of relaxed conditions, with only first-order conditions, and then with first- and second-order conditions.

### 3 Relaxation in the GNL model

#### 3.1 Derivation of Local Constraints

Now derive constraints on dissimilarity coefficients in the GNL model, as done in Herriges and Kling\(^1\). From condition 4 in GEV models, it is necessary to calculate mutual partial derivatives of choice probabilities for the derivation of constraints. Then, from equations (1)–(3), we can obtain the first- and second-order mutual derivatives as follow:

\[ \frac{\partial Q_n}{\partial V_{n'}} = \sum_{m=1}^{N_m} P_{mn}[m] P_{mn'}[m'] P_m \left( \tau_m - \frac{Q_{mn'}}{P_{mn}[m]} \right), \]  

\[ \frac{\partial^2 Q_n}{\partial V_{n'} \partial V_{n''}} = \sum_{m=1}^{N_m} P_{mn}[m] P_{mn'}[m'] P_{mn''}[m''] P_m \cdot \left( 2\tau_m^2 - \left( 1 + P_m + \frac{Q_{mn'}}{P_{mn}[m]} + \frac{Q_{mn''}}{P_{mn''}[m']} \right) \tau_m \right. 
\left. + \frac{2Q_{mn} Q_{mn'}}{P_{mn}[m] P_{mn'}[m]} - \sum_{m' \neq m} \tau_m P_{mn'} P_{mn'}[m'] P_{mn''}[m''] P_{mn''}[m'] \right). \]

Third order mutual derivatives are complicated, but can be found in the appendix.

Table 3.1 shows the upper limit of \( \mu_m \), conditional on the first-order conditions given choice probability \( Q_n \) and \( P_{mn}[m] \), based on equation (12). From Table 3.1, when \( Q_n \) and \( P_{mn}[m] \) are bigger, the upper limits of \( \mu_m \) are bigger. Therefore, if the alternative \( n \) belongs to the nest \( m \) and the utility of \( n \) is bigger, then the relaxation of \( \mu_j \) can be remarkable.

With relaxed parameters, the pseudo correlation in equation (7) cannot be interpreted. For example, consider the case \( \mu_m > 1, \alpha_{nm} \), and \( \alpha_{nm'} = 1 \). In this case pseudo correlation is negative, namely \( \rho_{mnm'} < 0 \). Note that pseudo correlation approximates precise correlation, and there are slight differences between them (Marzano and Papola\([11]\)). Unlike in the NL model, we must consider plural nests when we calculate the correlation between alternatives in the GNL model. This is because the operation \( \sum_{m=1}^{N_m} \) remains in condition (12), which is not in condition (10). After all, (12) is just elasticity of substitution.

On the other hand, relaxed parameters can be interpreted with psychology. The Hierarchical Elimination by Aspects (HEBA) model\([13]\) that is extended from the

---

\(^1\)For the case of the relaxation of allocation coefficients in the GNL model, see Takahashi and Ohno\([10]\).
Elimination by Aspects (EBA) model[14] is equivalent to the NL model (Batley and Daly [15]). There are no constraints on dissimilarity coefficients in the HEBA model that differ from those in the NL model. The GNL model is one of the Generalized Cross-Nested Logit models (Daly and Bierlaire [16]), so the relaxed dissimilarity coefficient in the GNL model \( \mu_m \) can be interpreted as the relative weight of a common aspect in the nest \( m \) to other aspects.

The NL model is derived from the GNL model with specifying parameters. Therefore, we verified the consistency of constraints between the GNL and NL models by deriving equations (12) and (13) from equations (10) and (11). We checked the consistency of local constraints in the GNL model with those in the NL model. The NL model can be derived from the GNL model as follows: \( \alpha_{nm} = 1, \alpha_{nm'} = 0 \). Then, each term in equation (12) and (13) becomes

\[
\frac{Q_n}{P_n|m} = P_m, \quad P_n|m = 0, \quad \forall m' \neq m,
\]

\[
\frac{\sum_{m' \neq m} \tau_{m'}P_n|m'P_{n'|m'}P_{m'}}{P_n|m} = 0.
\]

We can derive the first- and second-order conditions in the NL model from upper conditions as

\[
(12) \iff P_{n|m}P_{n'|m}P_m(\tau_m - P_m) \Rightarrow (10),
\]

\[
(13) \iff P_{n|m}P_{n'|m}P_{n'|m}P_m \cdot \{2\tau_m^2 - (1 + 2P_m + 3P_m^3 - 2P_m^2)\} = P_{n|m}P_{n'|m}P_{n|m}P_m \cdot \{2\tau_m^2 - (1 + 3P_m^3 - 2P_m^2)\} \Rightarrow (11).
\]

Third order conditions can be derived from the first- and second-order conditions. See the appendix.

### 3.2 Estimation Procedure

We need to establish an estimation procedure, a method calculating constraints, and an applied order of mutual partial derivatives. In this paper, we alternately need to conduct parameter estimations and fulfillment of constraints, owing to the dependence of constraints on estimated parameters in equations (12) and (13). Promising estimation candidates are the quasi-Newton and Markov Chain Monte Carlo (MCMC) methods. In the quasi-Newton method and other descent methods, the disadvantages are in reaching local maxima and the need to constrain at the initial setting. In the MCMC method, we constrain on parameters at initial settings moving with prior distributions as quasi-Newton methods, and it leads to longer calculation times.

This paper employs the Real-Coded Genetic Algorithm (GA), which is one of the metaheuristic algorithms. GAs are expected to reach global maximums, owing to parallel search, while keeping space of solution. It is easy to check for fulfilling constraints at each iteration in GA.

There are several methods for calculating constraints on dissimilarity coefficients that depend on the treatment of data and the calculation method of choice probabilities. This paper employs the following two methods. In the first method, we adopted the most rigid constraints among those derived from choice probability calculated in each revealed data (Most rigid). In the second method, we adopted averaged constraints from those derived from choice probability calculated in each revealed data (Average).

We used first- and second-order conditions at mutual derivatives, in order to compare to Kling and Herriges’ results. Kling and Herriges posited that it is sufficient for practical use to employ only first- and second-order conditions. However, they remarked that second-order conditions in the GNL model are not always more rigid than first order conditions. An alternative belongs to plural.
nests in the GNL model, where the derived condition in equation is not simple with respect to dissimilarity coefficients, where remains $\sum_{m=1}^{N_m} \theta_m$.

The estimation procedure via maximization of likelihood in this paper is shown in figure 1. In this figure, $\theta$ denotes the parameter set, the upper limit of constraints in nest $m$ is $\pi_m$, $L$ is log-likelihood, and superscription of $*$ means optimal value. In addition, superscriptions of $l$ and $k$ denote iteration numbers of Real-Coded GA and $\pi_m$, respectively.

At step 1, we conduct a log-likelihood estimation via Real-Coded GA under DZM conditions: $0 < \mu_m < 1 \forall m$, and obtained $\theta^*$. At step 2, we set the parameters obtained in step 1 as the initial values. At step 3, we calculate $\pi_m^k$ with given parameters and choice probabilities. In actuality, there is only one relaxed dissimilarity constraint. Then we calculate $\pi_m^L$ at only one nest. For steps 4 to 6, we estimate parameters via Real-Coded GA. In step 4, we generate the initial population. To guarantee the condition $L_l, k > L_{k-1, l}$, we employ $\theta_k$ for the part of the initial population. In step 5, we generate a new population. This paper employs simplex rules (SPX) for crossovers and random mutation to maintain diversity. In step 6, select from two parent chromosomes via reJGG, proposed in Iida et al.[17], which is extended from Just Generation Gap (JGG) proposed in Akimoto et al.[18]. In reJGG, parents that have high evaluation values tend not to be selected. The procedure for reJGG is as follows:

**Replication**: sampling $n_p$ times without replacement from the population, and set as parents for crossover.

**Generation of children**: apply crossover to parents repeatedly, and generate children for $n_c$ size.

**Selection**: select $n_p$ individuals from the children and parents who have top-$n_p$ evaluation values.

Repeat steps 5 and 6 while converging $L$ under given $\pi_m^L$ at inner cycle. Repeat steps 2 to 6 until the whole convergence of $L$, with renewal of constraints at outer cycle.

### 4 Application

This paper conducts a parameter estimation of the GNL model using ID-POS data at the two grocery stores in Japan used in Takahashi[12]. The target category is plastic bottled cokes for one year in 2001. The total number of point of sales is 5,269. The choice structure and aspects of each alternative are shown in Figure 2 and Table 2. In Figure 2, the upper and lower layers are the nest and alternative levels, respectively. Each number in the alternative level corresponds to the one in table 2. The number of parameters is 34.

| Brand   | Capacity (ml) | Sugar-Free or not |
|---------|---------------|-------------------|
| 1 CocaCola | 1,500        | Normal            |
| 2 CocaCola | 500          | Normal            |
| 3 CocaCola | 1,500        | Sugar-Free        |
| 4 CocaCola | 500          | Sugar-Free        |
| 5 Pepsi   | 1,500        | Normal            |
| 6 Pepsi   | 500          | Normal            |
| 7 Pepsi   | 1,500        | Sugar-Free        |
| 8 Pepsi   | 500          | Sugar-Free        |

For actual estimation, we set GA parameters as follows: the population $N_I = 340$, the number of crossovers $n_c = 340$, the mutation rate $p = 0.1\%$. The data are divided into two parts, for estimation and verification. We estimated using the most likelihood methods with data for estimation, and verified hit ratio or other statistics with data for verification.

Table 3 presents coefficients for five comparison models in this paper. In this table, first-order and second-order conditions are shown in (12) and (13), respectively. Table 4 provides the estimates under the Global, #1, and #2 Models. Table 5 shows the comparison result of estimated dissimilarity coefficients and goodness of fit under Global and #1–4 constraints. In Table 4, the dissimilarity coefficients in Models #1 and #2 are over the unit interval. If Models #1 and #2 involve the Global Model, then the log-likelihood of Models #1 and #2 become higher than that of the Global Model (Table 5). The difference of allocation coefficients among the #1, #2 and Global Models is significant, because there are tradeoffs among allocation coefficients. However, the difference of coefficients in the utility function among them is not significant. From Table 5, the relaxation of dissimilarity coefficients in models employing scheme 2 is more remarkable than in those employing scheme 1. In scheme 1, we calculated constraints on each consumer and employed the most rigid constraints. On the other hand, in scheme 2, if we calculated constraints with the average for all consumers,
1. CocaCola 2. Pepsi 3. 1,500ml 4. 500ml 5. Normal 6. Sugar-Free

![Diagram of choice structure](image)

Figure 2: Employed Choice Structure[12]

| Allocation Coefficients | Coefficients in Utility Function |
|-------------------------|----------------------------------|
| $\alpha_{11}$ 0.204 0.323 0.324 | $\beta_{1}$ $-0.006$ $-0.007$ $-0.007$ |
| $\alpha_{13}$ 0.630 0.497 0.495 | $\beta_{2}$ 0.370 0.468 0.470 |
| $\alpha_{15}$ 0.166 0.180 0.181 | $\beta_{3}$ 0.001 0.001 0.001 |
| $\alpha_{21}$ 0.424 0.508 0.508 | $\beta_{4}$ $-0.207$ $-0.270$ $-0.271$ |
| $\alpha_{24}$ 0.239 0.129 0.129 | $\beta_{5}$ 0.337 0.419 0.418 |
| $\alpha_{25}$ 0.337 0.362 0.363 | \text{Dissimilarity Coefficients} |
| $\alpha_{31}$ 0.079 0.119 0.119 | $\mu_{1}$ 0.200 0.319 0.320 |
| $\alpha_{33}$ 0.672 0.591 0.591 | $\mu_{2}$ 1.000 1.034 1.034 |
| $\alpha_{36}$ 0.249 0.289 0.290 | $\mu_{3}$ 0.204 0.217 0.216 |
| $\alpha_{41}$ 0.306 0.376 0.377 | $\mu_{4}$ 0.100 0.100 0.100 |
| $\alpha_{44}$ 0.197 0.118 0.118 | $\mu_{5}$ 0.100 0.100 0.100 |
| $\alpha_{46}$ 0.497 0.506 0.506 | $\mu_{6}$ 0.196 0.215 0.213 |

Table 4: Result of Estimates

then the impact of the most rigid consumers in scheme 2 is smaller than that of scheme 1. Relaxation of constraints leads to improvement in goodness of fit, but difference of estimation among schemes #1–4 does not make a significant difference in goodness of fit.

5 Conclusions

This paper derived local relaxed constraints on dissimilarity coefficients in the GNL model. For concrete, we derived first-order, second-order, and third-order derivative constraints based on Herriges and Kling. In this process, we designated mistakes of third-order derivative conditions in the GNL model proposed in Herriges and Kling, and revised the model. For reinforcement of our results, we verified equivalence between constraints in the NL and the GNL models. Furthermore, this paper proposed a parameter estimation method under obtained constraints and conducted comparison empirical analysis with ID-POS data. As a result, we showed higher log-likelihood and hit ratios under relaxed constraints than under DZM conditions. Its gain is slight, but the changes in allocation coefficients are significant.

Our task in the future is the derivation of higher-order conditions and the application of derived conditions to other nest structures. Constraints in the structure of the higher-level, such that Gil-Motló and Hole[19] are prospective to be different from the results in this paper.

ACKNOWLEDGMENTS

This study was supported by JSPS KAKEN ()

References

[1] C. H. Wen and F. S. Koppelman: The Generalized Nested Logit Model. Transp. Res. B, 63-2 (2001), 627–641.

[2] A. Daly and S. Zachary: Improved Multiple Choice Model. in Hensher, D. and Dalvi, Q. eds., Identifying and Measuring the Determinants of Mode Choice, (Saxon House, UK, 1979), 335–357.
Table 5: Relaxed Dissimilarity Coefficients and Goodness of Fit

|        | Global | #1     | #2     | #3     | #4     |
|--------|--------|--------|--------|--------|--------|
| $\mu_2$ | 1.000† | 1.034† | 1.034† | 1.152  | 1.153  |
| log-likelihood (estimate) | −5785.701 | −5779.594 | −5779.592 | −5779.447 | −5779.447 |
| log-likelihood (verification) | −2838.911 | −2835.601 | −2835.611 | −2835.538 | −2835.700 |
| AIC (estimate) | −11641.40 | −11629.19 | −11629.19 | −11628.89 | −11628.90 |
| AIC (verification) | −5747.822 | −5741.203 | −5741.221 | −5741.076 | −5741.400 |
| Hit Ratio (estimate)‡ | 40.94% | 41.03% | 41.03% | 41.00% | 41.00% |
| Hit Ratio (verification)‡ | 41.96% | 42.01% | 42.01% | 41.96% | 42.01% |

†: Values in under Global constraints, #1 and #2 are reproduced from Table 4.
‡: “Hit” means a consumer chose the one that had the highest estimated choice probability in the choice set.

[3] D. L. McFadden: Econometric Models of Probabilistic Choice. in Manski, C. F. and McFadden D. L. eds., Structural Analysis of Discrete Data (MIT Press, MA, 1981), 198–272.

[4] A. Börsch-Supan: On the Compatibility of Nested Logit Models with Utility Maximization. J. Econom., 43-3 (1990), 373–388.

[5] R. H. Koning and G. Ridder: On the Compatibility of Nested Logit Models with Utility Maximization: A Comment. J. Econom., 63-2 (1994), 389–396.

[6] C. L. Kling and J. A. Herriges: An Empirical Investigation of the Consistency of Nested Logit Models with Utility Maximization. American Journal of Agricultural Economics, 77-4 (1995), 875–884.

[7] J. A. Herriges and C. L. Kling: Testing the Consistency of Nested Logit Models with Utility Maximization. Econ. Lett., 50-1 (1996), 33–39.

[8] A. Papola: Some Developments on the Cross-Nested Logit Model. Transp. Res. B, 38-14 (2004), 833–851.

[9] D. L. McFadden: Modeling the Choice of Residential Location. In Karlqvist A., Lundqvist L., Snickars F. and Weibull J. eds., Spatial Interaction Theory and Planning Models, (Elsevier, Amsterdam, 1978), 75–96.

[10] K. Takahashi and T. Ohno: Equivalence between the GNL and Entropy Models: Equivalence in the Disaggregate Level. J. Jpn. Ind. Manage. Assoc., 64-1 (2013), 9–20, in Japanese.

[11] V. Marzano and A. Papola: On the Covariance Structure of the Cross-Nested Logit Model. Transp. Res. B, 42-2 (2008), 83–98.

[12] K. Takahashi: Fundamentals in the GNL models and Its Applications to Marketing. Doctoral Dissertation in Waseda Univ., (2013), in Japanese.

[13] A. Tversky and S. Sattath: Preference Trees. Psychol. Rev., 86-6 (1979), 542–573.

[14] A. Tversky: Elimination by Aspects: A Theory of Choice. Psychol. Rev., 79-4 (1972), 281–299.

[15] R. P. Batley and A. J. Daly: On the Equivalence between Elimination-by-Aspects and Generalised Extreme Value Models of Choice Behavior. Mathematical Psychol., 50-5 (2006), 456–467.

[16] A. J. Daly, and M. Bierlaire: A General and Operational Representation of Generalised Extreme Value Models. Trans. Res. B, 40-4 (2006), 285–305.

[17] Y. Iida, K. Takahashi and T. Ohno: A New Method for Parameter Estimation of the GNL Model Using Real-Coded GA. Operations Res. Proc. 2013, (2014), 209–215.

[18] Y. Akimoto, Y. Nagata, J. Sakuma, I. Ono and S. Kobayashi: Analysis of the Behavior of MGG and JGG as a Selection Model for Real-coded Genetic Algorithms. Trans. Jpn. Soc. Artificial Intell., 25-2 (2010), 281–289.

[19] M. J. Gil-Motló and A. R. Hole: Tests for the Consistency of Three-Level Nested Logit Models with Utility Maximization, Econ. Lett., 85-11 (2004), 133–137.

A Results of Higher Order

This appendix demonstrates the results of third-order mutual derivatives that were omitted in the main part.
Table 6: Partial Derivatives of Choice Probabilities in the NL and GNL models

| Model | Order | Derivatives |
|-------|-------|-------------|
| NL    | First | \( \frac{\partial Q_n}{\partial \mu_n} = P_{n|m} P_{n'|m} P_m \left( \frac{\mu_n - 1}{\mu_m} - P_m \right) \) |
| NL    | Second | \( \frac{\partial^2 Q_n}{\partial \mu_n^2} = P_{n|m} P_{n'|m} P_{n''|m} P_m \left( 2 \left( \frac{\mu_n - 1}{\mu_m} \right)^2 - (1 + 3P_m) \left( \frac{\mu_n - 1}{\mu_m} \right) + 2P_m^2 \right) \) |
| NL    | Third | \( \frac{\partial^3 Q_n}{\partial \mu_n^3} = P_{n|m} P_{n'|m} P_{n''|m} P_{n'''|m} P_m \left( 6 \left( \frac{\mu_n - 1}{\mu_m} \right)^3 - (7 + 11P_m) \left( \frac{\mu_n - 1}{\mu_m} \right)^2 + (12P_m^2 + 4P_m + 1) \left( \frac{\mu_n - 1}{\mu_m} \right) - 6P_m^3 \right) \) |
| GNL   | First | \( \frac{\partial Q_n}{\partial \mu_n} = \sum_{m=1}^{N_m} P_{n|m} P_{n'|m} P_{n''|m} P_{n'''|m} P_m \left( \frac{\mu_n - 1}{\mu_m} - \frac{Q_{n'} \mu_{n'}^2}{P_{n'|m} P_{n'''|m} P_m} \right) \) |
| GNL   | Second | \( \frac{\partial^2 Q_n}{\partial \mu_n^2} = \sum_{m=1}^{N_m} P_{n|m} P_{n'|m} P_{n''|m} P_{n'''|m} P_m \left( 2 \left( \frac{\mu_n - 1}{\mu_m} \right)^2 - \left( 1 + P_m + \frac{Q_{n'} \mu_{n'}^2}{P_{n'|m} P_{n'''|m} P_m} \right) \left( \frac{\mu_n - 1}{\mu_m} \right) + \frac{2Q_{n'} \mu_{n'}^3}{P_{n'|m} P_{n'''|m} P_m} \right) \) |
| GNL   | Third | \( \frac{\partial^3 Q_n}{\partial \mu_n^3} = \sum_{m=1}^{N_m} P_{n|m} P_{n'|m} P_{n''|m} P_{n'''|m} P_m \left( 6 \left( \frac{\mu_n - 1}{\mu_m} \right)^3 - \left( 7 + 5P_m + \frac{2Q_{n'} \mu_{n'}^2}{P_{n'|m} P_{n'''|m} P_m} + \frac{2Q_{n'} \mu_{n'}^3}{P_{n'|m} P_{n'''|m} P_m} \right) \left( \frac{\mu_n - 1}{\mu_m} \right) + \frac{2Q_{n'} \mu_{n'}^4}{P_{n'|m} P_{n'''|m} P_m} \right) \) |

International Journal of Japan Association for Management Systems