Tests for the consistency of three-level nested logit models with utility maximization.

María José Gil-Moltó and Arne Risa Hole*

Department of Economics, University of St Andrews.

December 10, 2003

Abstract

This paper provides necessary conditions for testing the local consistency of three-level nested logit models with random utility maximization. We find that for a model with two sub-nests per nest the conditions can lead to a substantial increase in the range of acceptable dissimilarity parameters, irrespective of the number of alternatives per sub-nest.

Keywords: Nested logit, Discrete choice, Random utility maximization

JEL classification: C25

*Corresponding author: Department of Economics, University of St Andrews, KY16 9AL, St Andrews, Scotland. Email address: arh12@st-and.ac.uk. Tel. no.: + 44 1334 462439; fax no.: + 44 1334 462444.

1 Introduction

The multinomial logit (MNL) model is the most widely used discrete choice model due to its closed-form choice probabilities and consistency with random utility maximization (RUM). However, the MNL model suffers from the restrictive independence from irrelevant alternatives (IIA) property which states that the ratio of two choice probabilities is independent of the other alternatives in the model. This implies that a change in an attribute of one alternative will have the same proportional impact on the probability of each of the other alternatives being chosen. The nested logit (NL) model relaxes the IIA property by dividing the alternatives into subsets or nests, allowing the IIA assumption to hold within each nest but not for alternatives in different nests. As opposed to the more flexible Multinomial Probit and Mixed Logit models, the NL model has closed-form choice probabilities which can be estimated without resorting to simulation methods. Due to its simplicity and allowing for a variety of substitution patterns, the NL model remains the most common extension of the MNL model in applied work.

Daly and Zachary (1979) and McFadden (1978a) have shown that the nested logit model is consistent with RUM under the condition that the dissimilarity parameters are constrained within the unit interval. In many practical applications, however, this condition has not been met. Börch-Supan (1990) argues that the DZM condition is unnecessarily strong given that the NL model should be viewed as a local approximation. Based on the work of Börch-Supan, Herriges and Kling (1996) derive necessary conditions for local consistency with utility maximization for two-level NL models.

The main contribution of the current paper is to extend the conditions of Herriges and Kling to be applicable to three-level nested logit models[1]. We develop explicit formulae that can be applied to test the consistency of three-level nested logit models with RUM. Since adding a level to the model implies that there are two sets of conditions that need to be satisfied (as opposed to one set in the two-level case) we pay particular attention to which combinations of dissimilarity parameters are acceptable. We find that for some nesting structures the conditions can lead to a substantial increase in the range of acceptable dissimilarity parameters.

2 The three-level nested logit model

Following Börch-Supan (1990), we assume a sample of T consumers with the choice of J discrete alternatives. The utility that individual t derives from choosing alternative j is denoted by u_{jt}. Utility is partitioned into a systematic component , v_{jt}, and a random component, ε_{jt}, such that:

$$u_{jt} = v_{jt} + \varepsilon_{jt} \tag{1}$$

The systematic component, v_{jt}, is a function of the attributes of alternative j and the individual's observable socio-demographic characteristics, while ε_{jt}

[1] Recent applications of the three-level nested logit model include Liaw and Frey (2003), Gabriel and Painter (2003), Shaw and Ozog (1999) and Eymann and Ronning (1997).

represents characteristics and attributes unknown to the researcher, measurement error and/or heterogeneity of tastes in the sample. Since the unknown variable, ε_{jt}, is treated as random by the researcher, this class of utility models is called random utility models. The probability that individual t chooses alternative i rather than alternative j is the probability that the utility of choosing i is higher than the utility of choosing j:

$$P_{it} = P(u_{it} > u_{jt}) = P(v_{it} + \varepsilon_{it} > v_{jt} + \varepsilon_{jt}) = P(\varepsilon_{jt} - \varepsilon_{it} < v_{it} - v_{jt}) \quad (2)$$

Denoting the joint density function of the random terms by, ε_t, the probability that alternative i is chosen is given by:

$$P_{it} = \int_{\varepsilon_t} I(\varepsilon_{jt} - \varepsilon_{it} < v_{it} - v_{jt} \forall j \neq i) f(\varepsilon_t) d\varepsilon_t \quad (3)$$

where $I(\cdot)$ is the indicator function, equalling 1 when the expression in parenthesis is true and 0 otherwise.

In the three-level nested logit model the alternatives are grouped in N subsets or nests, with $L(n)$ sub-nests in nest n and $J(l(n))$ alternatives in sub-nest $l(n)$. The choice can be visualized as first choosing among the N nests, then among the $L(n)$ alternatives in the chosen nest n, and finally among the $J(l(n))$ alternatives in the chosen sub-nest $l(n)$. Assuming that the joint density function of the random terms is given by a particular type of generalized extreme value (GEV) distribution (McFadden, 1978b), the probability of alternative $i(k(m))$ being chosen is given by (suppressing the individual subscript t

for simplicity):

$$P_i = P_{i|k(m)} P_{k|m} P_m \tag{4}$$

where $P_{i|k(m)}$ is the conditional probability of choosing alternative i given that sub-nest k and nest m are chosen, $P_{k|m}$ is the conditional probability of choosing sub-nest k given that nest m is chosen and P_m is the marginal probability of choosing nest m. The conditional and marginal probabilities are given by:

$$P_m = \frac{\exp(\mu_m IV_m)}{\sum_{n=1}^{N} \exp(\mu_n IV_n)} \tag{5}$$

$$P_{k|m} = \frac{\exp(\frac{\lambda_{k(m)}}{\mu_m} IV_{k(m)})}{\sum_{l=1}^{L(m)} \exp(\frac{\lambda_{l(m)}}{\mu_m} IV_{l(m)})} \tag{6}$$

$$P_{i|k(m)} = \frac{\exp(\frac{1}{\lambda_{k(m)}} v_{i(k(m))})}{\sum_{j=1}^{J(k(m))} \exp(\frac{1}{\lambda_{k(m)}} v_{j(k(m))})} \tag{7}$$

IV_m and $IV_{k(m)}$ are the inclusive values of nest m and sub-nest $k(m)$ respectively, where IV_m is given by the log of the denominator in (6) and $IV_{k(m)}$ by the log of the denominator in (7):

$$IV_m = \ln \sum_{l=1}^{L(m)} \exp(\frac{\lambda_{l(m)}}{\mu_m} IV_{l(m)}) \tag{8}$$

$$IV_{k(m)} = \ln \sum_{j=1}^{J(k(m))} \exp(\frac{1}{\lambda_{k(m)}} v_{j(k(m))}) \tag{9}$$

where the parameters μ_m and $\lambda_{k(m)}$ are the dissimilarity (or inclusive value) parameters for nest m and sub-nest $k(m)$ respectively.

3 Conditions for consistency with RUM.

McFadden (1981) has shown that any set of choice probabilities that satisfy the following compatibility conditions are consistent with RUM:

C.1 $P_j(v) \geq 0, \quad \sum_{j=1}^{J} P_j(v) = 1, \quad P_j(v) = P_j(v+\alpha), \quad \forall \alpha \in R$

where $v \equiv (v_1, ..., v_J)$, and

C.2 $\partial P_j(v)/\partial v_i = \partial P_i(v)/\partial v_j$

and finally,

C.3 P_j must have non-negative even and non-positive odd mixed partial derivatives with respect to components of v other than v_j.

In the case of the nested logit model only the final compatibility condition is restrictive. McFadden (1978a) and Daly and Zachary (1979) show that for C.3 to hold globally (for all $v \in R^J$) the dissimilarity parameters for the nests are restricted to lie within the unit interval:

$$\mu_n \leq 1, \quad \forall n \tag{10}$$

In addition the dissimilarity parameters for the sub-nests are restricted to be lower than or equal to their respective nest's dissimilarity parameter such that:

$$\lambda_{l(n)} \leq \mu_n, \quad \forall l(n) \tag{11}$$

Börch-Supan (1990) argues that the DZM conditions are unnecessarily strong given that the NL model should be viewed as a local approximation. His theorem 1 shows that if condition C.3 is satisfied for all observed and projected values of $v \in A$ where A is a subset of R^J, the probabilities are locally consistent with RUM. Intuitively, since economic theory restricts the region in which v is likely to lie, a sufficient condition for utility maximization is that C.3 holds for the values of v in this region. Furthermore, Börch-Supan's theorem 3 shows that for the NL model only derivatives of order less then the number of alternatives within each sub-nest and sub-nests within each nest need to be evaluated for C.3 to hold.

Building on Börch-Supan's results, Herriges and Kling (1996) derive necessary conditions for consistency with utility maximization for two-level NL models. In the case of the three-level NL model, Börch-Supan's theorem 3 implies that in addition to the conditions given by Herriges and Kling P_i, where $i \in l(n)$, must have non-negative even and non-positive odd mixed partial derivatives with respect to v_j for all $j \in l(n) \neq i(l(n))$. Thus we can derive the following theorem:

Theorem 1 *For three-level nested logit models, the following conditions are necessary for consistency with random utility maximization:*

$$\mu_n \leq \frac{1}{1-P_n}, \quad \forall n \tag{12}$$

$$\mu_n \leq \frac{4}{3(1-P_n)+[(1+7P_n)(1-P_n)]^{1/2}}, \quad \forall n \in G_3 \equiv \{n|L(n) \geq 3\} \tag{13}$$

$$\lambda_{l(n)} \leq \frac{1}{(1-P_{l|n})/\mu_n+(1-P_n)P_{l|n}}, \quad \forall l(n) \tag{14}$$

$$\lambda_{l(n)} \leq \frac{4}{3/\mu_n+3P_{l|n}-3(1/\mu_n+P_n)P_{l|n}+F^{1/2}}, \quad \forall l(n) \in S_3 \equiv \{l(n)|J(l(n)) \geq 3\} \tag{15}$$

where,

$$F = (1+7P_n)(1-P_n)P_{l|n}^2+(1+7P_{l|n})(1-P_{l|n})/\mu_n^2-6/\mu_n(1-P_n)(1-P_{l|n})P_{l|n}$$

The proof follows from differentiation of equation (4) and is available from the authors upon request. Equations (12) and (13) correspond to the conditions in Herriges and Kling, while (14) and (15)[2] are implied by the first and second

[2] We have verified that $F > 0$ for all $\mu_n \leq 1/(1-P_n)$.

order mixed derivatives of P_i with respect to v_j for $j \in l(n) \neq i(l(n))$.

The conditions in theorem 1 can be used to test *ex post* the local consistency of three-level NL models with RUM. Alternatively they can be used as a guide for specifying Bayesian priors for the dissimilarity parameters (see Lahiri and Gao, 2002). The conditions are necessary and sufficient for a model with three alternatives per sub-nest and three sub-nests per nest. For a model with two alternatives per sub-nest and two sub-nests per nest (12) and (14) are necessary and sufficient. The conditions are not sufficient when there are more than three alternatives per sub-nest, but in practical applications testing the first and second-order conditions may be considered satisfactory (see Kling and Herriges, 1995 for a discussion).

It can be seen that substituting $P_n = 0$ into (14) or (15) and $P_n = P_{l|n} = 0$ into (16) or (17) yields the DZM conditions. For $P_n > 0$ the upper limit for μ_n exceeds unity as shown by Börch-Supan and Herriges and Kling. Also, for values of $P_{l|n}$ higher than 0, $\lambda_{l(n)}$ is no longer restricted to be lower than or equal to μ_n. In other words the range of acceptable combinations of μ_n and $\lambda_{l(n)}$ can be expanded beyond the bounds imposed by equations (12) and (13). The rest of our paper is devoted to investigating to what extent the conditions in theorem 1 expand the acceptable combinations of μ_n and $\lambda_{l(n)}$ given values for P_n and $P_{l|n}$ and different nesting structures.

Figure 1 plots the upper bounds on $\lambda_{l(n)}$ as a function of μ_n given different values for the probabilities. Lines 1 and 2 represent the first and second order restrictions on $\lambda_{l(n)}$ respectively (eqs. 14 and 15), while 3 and 4 represent the first and second order restrictions on μ_n (eqs. 12 and 13). The area enclosed by lines 1/2 and 3/4 represent acceptable combinations of $\lambda_{l(n)}$ and μ_n given different nesting structures. These can be compared to the shaded area, which gives the acceptable combinations of $\lambda_{l(n)}$ and μ_n given that the DZM conditions are imposed. As shown by Herriges and Kling, the upper bound on μ_n decreases markedly when the number of sub-nests per nest grow. Also, the upper bound on μ_n increases in P_n. Increasing the number of alternatives in each sub-nest, however, does not lead to a substantial decrease in the acceptable values for $\lambda_{l(n)}$ when $P_{l|n}$ is low. Even for higher values of $P_{l|n}$ the range of acceptable combinations of the dissimilarity parameters is considerably greater than in the DZM case given that the number of sub-nest per nest is not higher than two and P_n is high.

[Insert figure 1 near here]

4 Concluding remarks

We have developed conditions that can be used to test the local consistency of three-level nested logit models with random utility maximization. We find that irrespective of the number of alternatives per sub-nest the conditions can lead to a substantial increase in the range of acceptable dissimilarity parameters for a model with only two sub-nests per nest. Since this model structure is not uncommon in the literature the result is of relevance to practitioners who find that the estimated dissimilarity parameters fall outside the unit range.

References

[1] Börch-Supan, A., 1990, On the compatibility of nested logit models with utility maximization, Journal of Econometrics 43, 373-388.

[2] Daly, A. and S. Zachary, 1979, Improved multiple choice models, in: D. Hensher and Q. Dalvi, eds., Identifying and measuring the determinants of mode choice (Teakfield, London) 335-357.

[3] Eymann, A. and G. Ronning, 1997, Microeconometric models of tourists' destination choice, Regional Science and Urban Economics 27, 735-761.

[4] Gabriel, S. and G. Painter, 2003, Intra-metropolitan mobility, residential location, and homeownership choice among minority and white households: estimates of a nested multinomial logit model, Paper presented at the Annual Meetings of the American Real Estate and Urban Economics Association, Washington D.C.

[5] Lahiri, K. and J. Gao, 2002, Bayesian analysis of nested logit model by Markov chain Monte Carlo, Journal of Econometrics 111, 103-133.

[6] Kling, C.L. and J.A. Herriges, 1995, An empirical investigation of the consistency of nested logit models with utility maximization, American Journal of Agricultural Economics 77, 875-884.

[7] Herriges, J.A. and C.L. Kling, 1996, Testing the consistency of nested logit models with utility maximization, Economics Letters 50, 33-39.

[8] Liaw, K.-L. and William H.F., 2003, Location of Adult Children as an Attraction for Black and White Elderly 'Return' and 'Onward' Migrants in the United States: Application of a Three-Level Nested Logit Model with Census Data, Mathematical Population Studies 10, 75-98.

[9] McFadden, D., 1978a, Quantitative methods for analyzing travel behaviour of individuals: Some recent developments, in: D. Hensher and P. Stopher, eds., Behavioral Travel Modelling (Croom Helm, London) 279-318.

[10] McFadden, D., 1978b, Modelling the choice of residential location, in: A. Karlquist et. al., eds., Spatial interaction theory and residential location, (North-Holland, Amsterdam) 75-96.

[11] McFadden, D., 1981, Econometric models of probabilistic choice, in: C.F. Manski and D. McFadden, eds., Structural analysis of discrete data with econometric applications (MIT Press, Cambridge, MA).

[12] Shaw, D.W. and M.T. Ozog, 1999, Modeling overnight recreation trip choice: application of a repeated nested multinomial logit model, Environmental and Resource Economics 13, 397-414.

Figure 1. Upper bounds on $\lambda_{l(n)}$ as a function of μ_n for different probability values.