The Journal of Choice Modelling ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

The Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Variance-component-based nested logit specifications: Improved formulation, and practical microsimulation of random disturbance terms

David S. Bunch [a,*], David M. Rocke [b]

[a] Graduate School Management, University of California, Davis, USA
[b] Division of Biostatistics and Department of Biomedical Engineering, University of California, Davis, USA

ARTICLE INFO

Article history:
Received 4 October 2015
Received in revised form
2 March 2016
Accepted 5 April 2016

Keywords:
Discrete choice modeling
Nested logit
Random utility maximization
Monte Carlo simulation
Random disturbance terms

ABSTRACT

The initial motivation leading to the results in this paper was a problem most choice modeling researchers may have not considered: how to simulate random disturbance terms from nested logit (NL) models. We develop an approach using results from Cardell (1997), who proved the existence of a probability distribution ($C(\lambda)$) that could be used to formulate NL models based on statistically independent variance components. These components can be interpreted as unobserved preference heterogeneity for the choice 'dimensions' used to define NL tree structures. Simulation aside, we consider this formulation to have other practical advantages for empirical work, but it does not appear to have penetrated the literature (possibly due to notational obstacles). We use notation from Daly (2001) to implement an equivalent representation, which also establishes mathematical equivalence between Cardell (1997) and other important results in the NL literature.

1. Introduction

To provide context, consider a simple discrete choice example (mode choice for commute trip) expressed in the random utility maximization (RUM) framework:

$$U_j = V_j^* + \varepsilon_j = \alpha_j + X_j\beta + \varepsilon_j,\ j \in \{drive\ alone, carpool, bus, subway\}. \tag{1}$$

where α_j's are alternative-specific constants (ASCs), X_j is a $1 \times p$ vector of explanatory variables, β is a p-vector of parameters, and ε_j's are random disturbance terms.[1] If the ε_j's are *iid* Gumbel (type I extreme value) with scale μ, then the probability that alternative i is chosen $\left(Prob\left\{U_i \geq U_j, \forall\, j \in C\right\}\right)$ is given by the multinomial logit (MNL) model

$$P_i = \frac{e^{\mu V_i^*}}{\sum_{j \in C} e^{\mu V_j^*}} \tag{2}$$

* Corresponding author.

[1] The vector X_j can be a function of consumer-specific characteristics in addition to attributes of choices alternatives, an aspect suppressed in this notation. The "fixed" utility V_j^* includes a star for notational reasons arising later in the paper.

http://dx.doi.org/10.1016/j.jocm.2016.04.001
1755-5345/

where an important early reference is McFadden (1973).[2] The MNL's well-known limitations (e.g., independence from irrelevant alternatives, unrealistic substitution patterns across alternatives, etc.) are due to the *iid* assumption, and many of the methodological advances in discrete choice modeling can be viewed as efforts to relax this assumption. These generally fall into two categories: simulation-based models (e.g., mixed MNL, a.k.a., mixed logit, multinomial probit), and generalized extreme value (GEV)-based models. At the risk of oversimplifying: the first approach allows more flexibility for implementing behavioral concepts (e.g., individual-level taste variation, especially in repeated measures/panel data), but requires simulation methodologies for estimation and analysis, whereas the second approach yields closed-form expressions for models with non-independent correlation structure (which avoids simulation), but at the cost of more complexity. Either approach has its own practical challenges, e.g., identification issues, and the need for specialized software. For additional background, see Train (2009).

This paper focuses on nested logit (NL) models (from the GEV family), and was initially motivated by an unusual need arising from a research project on energy systems models: simulating random disturbance terms from a pre-existing, complex NL model[3]. NL modeling needs are almost universally limited to computing choice probabilities (for model estimation and/or prediction) which can be done using closed-form expressions, so this particular problem has received scant attention in the literature (see discussion in Section 5).

The approach we developed uses results from Cardell (1997) ["Cardell"], who *(i)* shows how the total disturbance term in (1) can be decomposed into independent variance components, and *(ii)* establishes the existence and properties of probability distributions for these components. This high-level description suggests an obvious solution: simulate the total disturbance term by generating a random draw for each component, and then add them together. Although this may sound straightforward, there are a number of technical challenges, all of which are arguably a result of the model's mathematical complexity. For example, although fundamental theoretical results for NL were (independently) established by Williams (1977), and Daly and Zachary (1978), and subsequently generalized by McFadden (1978, 1981) to the GEV family, confusion over a variety of issues eventually led to controversies in the literature in the late 1990s and early 2000s.

Carrasco and Juan De Dios Ortúzar (2002) provides a review, and addresses these in a rigorous and concise way. They demonstrate how two approaches with alternative interpretations ("Williams/Daly-Zachary" and "McFadden") lead to mathematically equivalent NL formulations, and review important implementation issues that require researchers to make choices among specification and normalization options that can have important implications for the model's properties. Their review is framed in the context of two-level NL models, using a widely used form of notation. In this paper we use a more general form from Daly (2001) ["Daly"], for reasons to be discussed. However, because their treatment is general, we frequently refer to "Carrasco-Ortúzar" to avoid unnecessary replication of technical detail. We combine Daly's notation with Cardell's results to derive expressions that support a practical solution to the simulation problem. They also establish mathematical equivalence among various model forms and derivations in the literature (which is necessary for them to be truly useful). In this regard, we view this as a natural extension of Carrasco-Ortúzar, with additional practical advantages for empirical work.

Section 2 establishes required NL notation and definitions. Section 3 summarizes relevant results from Cardell, and provides expressions using Daly notation that establish key equivalences. Section 4 describes generation of random variates for variance components, and Section 5 concludes with comments on the approach, and related results in the literature.

2. Nested logit notation and normalization

First, recall the mode choice problem defined in (1). This problem appears frequently in the literature, where one possible NL tree structure is in Fig. 1. It depicts a hierarchy based on an upper-level "choice dimension" defined by nominal features (Auto versus Transit) that are shared by subsets of lower-level alternatives. Such multi-dimensional structure is common in transport applications (e.g., destination-mode choice, or destination-route choice), but, more generally, tree structures provide a useful way to represent perceived "similarity" or "substitutability" among competing alternatives, and how they might vary as a function of such shared choice dimensions. In the context of random utility maximization, this can also be interpreted in terms of the degree of correlation among the random disturbance terms in (1). As discussed in Carrasco-Ortúzar (and also here), there are multiple pathways for deriving NL models based on alternative conceptual interpretations that yield mathematically equivalent formulations.

As previously noted, we use notation from Daly (2001). Let c denote any node in the tree, which includes *elemental alternatives* (that can also be denoted by e) and a root node (*root*). The *tree function* $t(c)$ identifies the (unique) parent (precursor) of c. The set of c and its ancestors is $A(c) = \{c, t(c), t(c(c)),\ldots, t(c)=k \mid t(k) = root\}$.

Daly distinguishes between *normalized* and *non-normalized* forms. We are using the normalized form, where each non-elemental node c (a *composite node* that subsumes a subset of elementary alternatives) is assigned a structural parameter μ_c. Models are recursively defined using functions V that represent the 'attractiveness' of nodes, defined for elementary and

[2] The CDF takes the form $F_\varepsilon(y) = \exp(-e^{-\mu y})$

[3] Specifically, we developed an approach for incorporating heterogeneous consumer preferences for alternative fuel vehicles within a linear programming-based energy systems model. This required generating realizations of random utilities from a NL model that included 40 choice alternatives and up to four "levels" in the tree: for details, see Bunch et al. (2015).

Please cite this article as: Bunch, D.S., Rocke, D.M., Variance-component-based nested logit specifications: Improved formulation, and practical microsimulation of random disturbance terms. The Journal of Choice Modelling (2016), http://dx.doi.org/10.1016/j.jocm.2016.04.001