Properties, Advantages, and Drawbacks of the Block Logit Model

Jeffrey Newman

Michel Bierlaire

STRC 2009 **September 2009**

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Abstract

This paper proposes a block logit (BL) model, which is an alternative approach to incorporating covariance between the random utilities of alternatives into a GEV random utility maximization model. The BL is similar to the nested logit (NL), in that it is a restricted form of a network GEV (NGEV) model. The NL is a NGEV model where all of the allocation parameters have been fixed with a value equal to zero or one. The BL model proposed in this paper imposes the other possible constraint on the parameters of a NGEV model, so that the restrictions are placed not on the allocation parameters, but on the logsum parameters.

The BL model can be used in much the same way as the NL model, to generate choice models that exhibit inter-alternative correlations in random utilities. It can reproduce similar covariance structures to those generated by NL models, but can also create a wider variety of possible covariance matrices than the NL, because it allows overlapping blocks.

The estimation of parameters for the BL model is difficult, as the likelihood function for BL models is non-continuous. This paper examines the shape of the BL log likelihood function in some detail, and compares the relative performance of estimation procedures for NL and BL models using mode choice data.

1 Introduction

Research on generalized extreme value (GEV) discrete choice models has generally tended towards ever more general and complex forms. The process began with the introduction of the GEV formulation by McFadden in 1978, where he laid out the general framework for generalized extreme value models, and demonstrated that the multinomial logit (MNL) and nested logit (NL) models were consistent with that framework. Since the development of the GEV structure, various new forms of GEV model have been introduced, exhibiting more varied covariance structures. Such models as the paired combinatorial logit (Chu, 1989), the cross-nested logit (Vovsha, 1997), the ordered GEV (Small, 1987), and the product differentiation model (Bresnahan *et al.*, 1997), and the generalized nested logit (Wen and Koppelman, 2001) model provided various flavors of GEV for modelers to use. Each successive iteration of the models was typically more general than the previous, allowing for more parameters to articulate a wide array of potential multivariate error structures. The most recent iterations of this process has been in the network GEV model (NGEV) (Daly and Bierlaire, 2006), and the heterogeneous NGEV model (Newman, 2008a). These models are the most general of all of these models, with a plethora of parameters and structural forms that allow myriad possibilities for error distributions.

Yet, as with a marionette that has too many joints and strings, the increased flexibility of the most advanced models may not be worth their complexity, as they can contort and tangle in a variety of cumbersome ways, and only the most skilled puppeteers or modelers will be able to use them in any reasonable fashion. In many cases, it is desirable to use a more simplified models that the NGEV, so as to limit the number of parameters, and ensure a more stable, if less tailored, fit.

This paper examines model simplification from a historically backwards viewpoint: instead of working from the simpler MNL and NL models up to more general and flexible structures, we start with one of the the most recent and general structures, the NGEV model, and look at ways to distill it to a simpler, less flexible versions. One such distillation returns down the historical path, restricting allocations and eliminating cross-overs to return the the NL model form. However, an alternative distillation removes the estimation of logsum parameters, and controlling corellation only through blocks of allocations to nests. We call the resulting form a "block logit" (BL) model, and in this paper we will lay out its mathematical formulation, examine its properties, advantages, and drawbacks, and look at a sample application of the model on mode choice data. While the BL model is certainly not appropriate for use in many applications, it has some unusual properties that may make it appealing in particular situations. And even for modelers who have no intention of employing this unusual tool, this an analysis of the attributes and behavior of the BL model can inform a modeler's understanding of other similar models, including the CNL and NGEV models.

2 Model Definitions

2.1 Network GEV

A generalized extreme value model, as defined by McFadden (1978), creates a discrete choice model from a generating function $G(y)$ that is consistent with a few rules:

1. $G(y_1, y_2, ..., y_J)$ is a non-negative, homogeneous of degree one[1] function of $(y_1, y_2, ..., y_J) \geq 0$
2. $\lim_{y_i \to +\infty} G(y_1, y_2, ..., y_J) = +\infty$
3. The partial derivatives of G with respect to any distinct $i_1, i_2, ..., i_k$ from $y_1, y_2, ..., y_J$ are non-negative for odd k and non-positive for even k.

The NGEV model Daly and Bierlaire (2006), as discussed earlier, is one of the most general GEV models. It is defined by a finite, circuit-free, directed network, with a single source or root node, a sink node for each elemental alternative, and a set of GEV generating functions for each node i in said network (Newman, 2008b):

$$G^i(y) = \left(\sum_{j \in i^{\downarrow}} \left[\left(\alpha_{ij} G^j(y) \right)^{1/\mu_i} \right] \right)^{\mu_i}, \quad (1)$$

where $i^{\downarrow}$ is the set of successor nodes of i in the network, and defining $G^j(y) = y_j$ for the elemental alternative nodes. The generating function for the network's root node is a complete GEV compliant generating function for the network GEV model, so long as $\alpha \geq 0$, some $\alpha_{ij} > 0$ for all j, and $\mu_i > \mu_j$ for all j in $i^{\downarrow}$. Beyond any parameters embedded in the alternative utility functions inside y, the NGEV model is parameterized by a set of logsum parameters μ (one assigned to each network node), and a set of allocation parameters α (one assigned to each network arc).

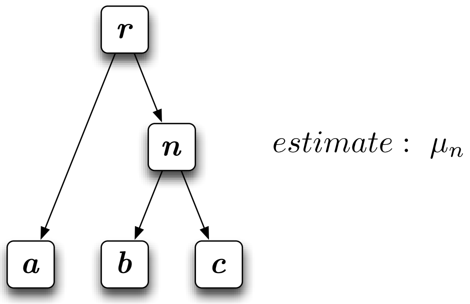
2.2 Nested Logit

The nested logit model has been in use for many years, and will be familiar to most readers. It is a specific case of the network GEV model, where for each set of arcs entering a node, exactly one arc has $\alpha = 1$, and all others have $\alpha = 0$. As such, it is defined as a GEV model using

$$G(y) = \sum_n \left(\sum_{j \in n} \left[y_j^{1/\mu_n} \right] \right)^{\mu_n}$$

[1] Ben-Akiva and François (1983) generalized this to homogeneity of any positive degree, but the degree-one condition is suitable here.

Figure 1: Nested Logit Model, Schematic Form



as a generating function when there is one level of nests, and

$$G(y) = \sum_{n} \left(\sum_{j \in n} \left[\left(\sum_{k \in j} y_j^{1/\mu_j} \right)^{\mu_j/\mu_n} \right] \right)^{\mu_n}$$

when there are two levels of nests, and expanding the formula recursively for any number of hierarchical levels. In the most simple form, the model relates the extreme value distributed error terms for the utilities of three alternatives, two of which are nested together, leaving the remaining alternative separate from the nest. A schematic network for such a model is depicted in Figure 1.

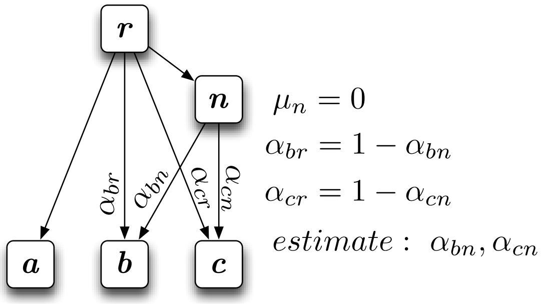
2.3 Block Logit

The imposition of values on the α parameters in a NGEV model to create a NL model leads to a Frostian exercise: what sort of model emerges if the restrictions are reversed? That is, the μ parameters are restricted to binary values (as for α, the binary values are the extremes of their valid interpretable values in a typically normalized model), and the α parameters are those to be estimated. In particular, the μ parameter for the root node is constrained to be 1, and for all other nesting node, the parameter is constrained to be 0. Such a form results in what we will call a "block logit" (BL) model.

Actually setting $\mu = 0$ directly is problematic, as the exponents in (1) include both μ and $1/\mu$, creating an undefined value. Instead, we can incorporate this extreme value for μ by using the limit of the generating function as μ approaches zero, giving

$$\begin{aligned} G(y) &= \sum_{j \in C} \alpha_{jr} y_j + \sum_{n \in N} \left(\lim_{\mu \to 0} \left(\sum_{i \in n} \left[(\alpha_{in} y_i)^{1/\mu} \right] \right)^{\mu} \right) \\ &= \sum_{j \in C} \alpha_{jr} y_j + \sum_{n \in N} \left(\max_{i \in n} (\alpha_{in} y_i) \right), \qquad (2) \end{aligned}$$

Figure 2: Block Logit Model, Schematic Form



with $\alpha_{in} > 0$ for at least one n (a nest or the root) for all i, and $\alpha_{in} \geq 0$ for all i and n. This is essentially the traditional most general form of the cross-nested logit model, as in Bierlaire (2006), but with two modifications. First, in addition to the nests, there is one additional connection directly from each elemental alternative to the root node of the model. And secondly, of course, is the limit formulation for μ, which imbues the model with some unique properties. The restriction that some α must be strictly greater than zero for each j is necessary to ensure that in all cases the BL model remains compliant with the conditions for a GEV model.

Schematically, the BL model in its simplest form looks similar to the NL model, with three alternatives, two of which are paired in a nest. However, as shown in Figure 2, each of the nested alternatives is also connected to the root directly. There are allocation parameters that control the split between the nest and the root for all nodes in the nest, although for identification purposes not all of the allocation parameters can be estimated. Instead, only one allocation parameter is estimated for each elemental alternative, and the other is set such that the total allocation of each alternative adds to one.

By including the root-linking first term in (2), and in particular if all $\alpha_{ir} > 0$, the BL retains one useful feature of most other GEV models, that the probability of every alternative is strictly greater than zero. Without the root-linking term, the BL model becomes a hybrid logit/best response model, where each nest represents a deterministic best-response calculation among a particular subset of choices, and only the best responses in each subset are considered within a probabilistic logit framework. In this case, it is very easy for individual alternatives to have a probability exactly equal to zero. This can be a problem if any such alternative is chosen in an observation, as the log likelihood for the entire model in that case would be negative infinity, no matter how well the model fit any number of other observations. This can be inconvenient, but is not necessarily a fatal flaw for this model form. Indeed, the non-gradient estimation procedures required, as will be described in section 4.1, should in most instances be able to handle this problem. It is conceivable that a pathological case could combine data and a BL model structure so as to make the solution for any finite log likelihood infeasible, but this would indicate a poor choice of the particular modeling structure, and could likely be overcome with a better structure.

The BL model also collapses any hierarchical structure of nesting blocks into a flat (single level) structure, with no nest containing another nest. This is a result of the NGEV's collapsability feature of sequences of nests that share identical logsum parameters, and is a particularly obvious result flowing from (2), where the maximization operator merges any possible subnests together, with only one alternative emerging as relevant. To paraphrase the rules of *Thunderdome*: n alternatives enter, one alternative leaves.

3 The shape of likelihood

A consideration of the differences between the block and nested logit models is helped by examining the shape of the log likelihood function for a very simple case.

Consider a single decision maker and a single choice among three alternatives, A, B, and C. Those alternatives have explanatory variables $x_A = -1$, $x_B = 0$, and $x_C = 1$, and an underlying relationship between alternatives A and B, separate from the observed data. This test case is illustrative but reasonably general. If the observed data for A and B were identical, then their relative probabilities would be equal in all cases, and not especially enlightening. The value for C is relatively unimportant, as its behavior vis-à-vis A and B is governed by an MNL type relationship.

To model this test case, the relationship between A and B can be incorporated in a traditional nested logit model, which results in two model parameters, a β parameter on the explanatory variables, and a μ parameter to describe the unexplained relationship. Alternatively, this scenario could be modeled using a block logit model, using one of two approaches. The model could have two parameters that would mimic the NL form: again a β parameter on the explanatory variables, and a single α parameter to describe the unexplained relationship. Or, the model could employ three parameters, β as well as a separate α for each alternative; this situation will be addressed separately, below.

The log likelihood function is calculated as a result of the actually selected alternative, either A, B, or C. If the decision maker chose alternative A, the log likelihood functions for these two models would take the forms illustrated in Figure 3. The two functions share some common features. The field of signs of the derivatives of the surface are the same, and the front and back sides of the functions (where μ and α are both 0 or 1, respectively) are identical. Between those identical sides, however, is an obvious difference. The block logit surface provides a log-linear interpolation between the front and back sides, resulting in a discontinuity (i.e., a cliff) in the log likelihood function. The nested logit surface, on the other hand, provides a smoothed transition between the front and back sides, with the discontinuity prevented as long as $\mu > 0$.

Figure 3: Comparative Log Likelihood of Alternative A

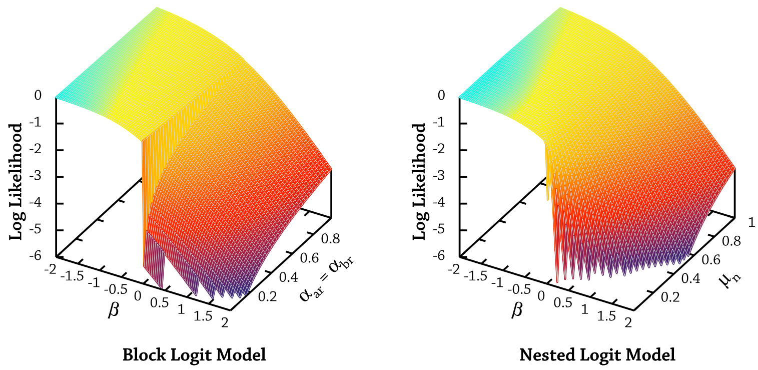
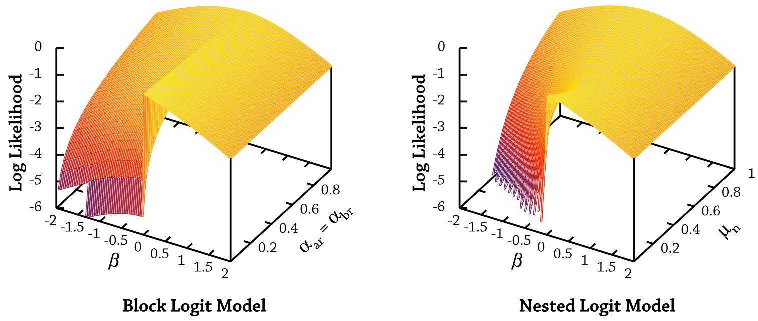


Figure 4: Comparative Log Likelihood of Alternative B



It is clear that in estimation the blocked model, with its discontinuous surface, could pose some problems for traditional gradient-based maximum likelihood algorithms if employed directly. While the single-case log likelihood function depicted here has only a single cliff that is consistant with the gradient in its region, estimating a model with multiple cases and multiple β parameters can easily result in numerous discontinuities, in different places, and which are not all aligned with the localized gradients. The nested model, which is smooth, functions better with those algorithms. In practice, however, this difference may or may not be all that important. The BL model can be closely approximated by a CNL model where the logsum parameters are constrained to be a suitably small value. In this case, the basic discontinuous nature of the BL model is mitigated and gradient-search optimization techniques can be applied. The resulting log likelihood function will by quite lumpy when applied to datasets with multiple observations, and there are likely to be very many local optima; modern heuristics for finding global optimum values should generally be applied in this case. Yet this criticism can apply as well to nested logit and other GEV models beyond the MNL model (which is globally concave).

If the decision maker chose alternative B, the log likelihood functions for these two models would take the forms illustrated in Figure 4. The shape of the functions is different than for A, although the discontinuity issues are the same. For this case, the global optimum for the both models is on the front edge of the surface depicted in the figure. However, for the NL model, that optimum is a small peak surrounded by a region of lower likelihood, which would be more difficult (though certainly not impossible) for a gradient search algorithm to find, especially if using a starting point on the rear side of the surface, near where the MNL optimum would be. The BL model, on the other hand, has a ridge extending from the MNL optimum towards to global optimum. Of course, such a simple case with only one observation provides little insight into the real relative performance of optimization in NL and BL models, but it does highlight some of the local-optimum risks of which modelers should be aware.

The form of the model does not notably change the shape of the log likelihood function if the decision maker chose alternative C, so those are not shown.

3.1 Seperate α values

As mentioned above, the BL model provides the opportunity of estimating seperate α values for each alternative. Within the simple one-case framework presented here, that would result in two parameters for controlling the relationship between A and B, instead of just one. If the modeler is interested in just creating a corellation between the alternatives, using multiple parameters to do so is overkill, as the same corellation can be obtained through many different combinations of α values. However, such a model is not mathmatically overspecified, as the different combinations of α parameters will create joint error distributions which do differ, but only in the third or higher moments.

Figures 5 and 6 show the basic log likelihood function for alternative A in the test model, holding each α parameter constant at various values. In Figure 5, α_{ar} is held constants at various values, and as a result of the BL formulation a curved slice of the log likelihood function is penalized, for sending a portion of A into competition with a portion of B and losing. The missing bite from the last surface of Figure 5 is a log likelihood of $-\infty$, because α_{ar} is 0, forcing all of A into a best-response block with B, which in the missing bite region it loses. The surfaces in Figure 6 are generated by holding the other α parameter constant, and also result in a curved penalty region, however the curve is in a different direction, and the penalty shape is different. In these surfaces, the benefit of winning the block is more clearly visible, as the shrinking “win” region is increasingly green.

Figure 5: Likelihood of Alternative A, given fixed values of α_{ar}

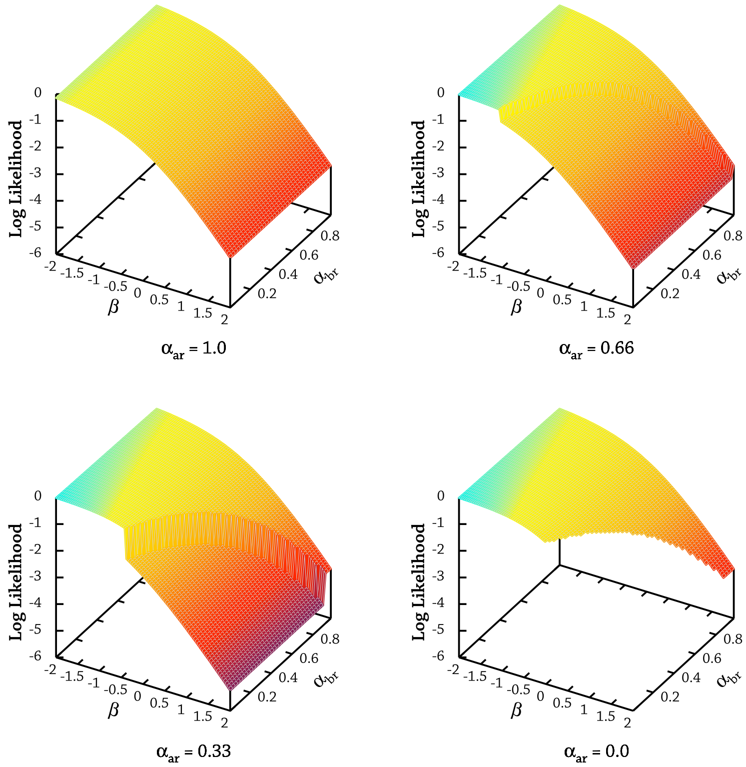
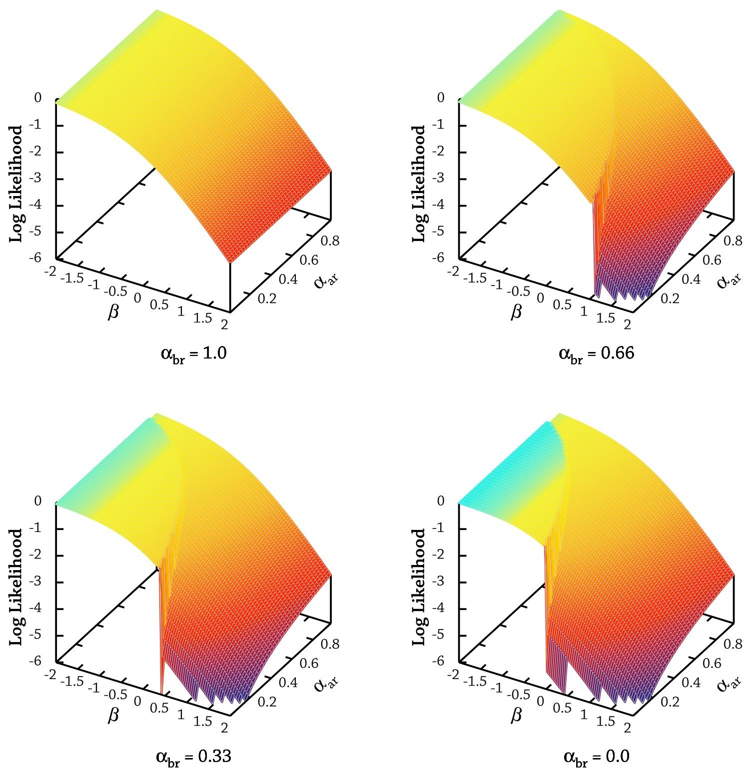


Figure 6: Likelihood of Alternative A, given fixed values of α_{br}



4 Pros and Cons

4.1 Disadvantages in using the block logit model

The BL model is substantially more difficult to work with than the NL model, as has been suggested by the review of the form of the log likelihood function. A primary reason for this difficulty arises from the non-continuous nature of the log likelihood function. Estimating parameters for the BL model directly requires non-gradient methods, and will in many circumstances be much slower than the NL model. An alternative approach is to approximate the BL model with a CNL model that constrains the μ parameters to a suitably small value. This can formally smooth over the cliffs in the BL model, but may leave a rugged landscape in the likelihood function that will still be difficult to search for a global optimum.

This discontinuity problem, though, highlights an important issue with the CNL model, and to a lesser extent, the NL model as well. The discontinuity arises due to the extreme value of the logsum parameters (i.e. they are equal to zero). The functional form for the CNL and NL models, where the logsum parameters are not zero, approaches this discontinuous shape as the logsum parameters approach zero. While gradient-based searches will function "correctly" for the CNL and NL models, the results of such local searches will be highly suspect, and more advanced global optimum seeking heuristics should always be used for these models.

The BL model also will have more parameters to estimate than the NL model, for a similar alternative-correlation form. The BL has one allocation parameter for each alternative-nest combination, while the NL has only one parameter for each nest. In most conceivable practical applications of the BL model, many of these parameters will be constrained to have identical or related values, so that the number of estimated parameters would be similar, but those relationships must be carefully defined.

If some connection between allocation parameters is not enforced, the relationships between alternatives can be difficult to interpret. If the allocation parameters are not symmetric among alternatives, the resulting covariance structure may not be apparent to a modeler. It is also possible to specify many different sets of allocation parameters that will generate different choice models, but which have the same covariance structure; the random utility distributions will differ only in their third or higher moments. These differences, while real, are very subtle, and modelers will have very little statistical or intuitive guidance as to selecting the best model.

4.2 Advantages in using the block logit model

The BL does have a few positive traits that mitigate these problem: a correlation structure that is more flexible than that of the NL, and possibly a simpler application of the model for forecasting, particularly when the number of possible alternatives is large.

The heirarchical nesting structure of the NL model can be very restrictive when creating models, as one alternative cannot be simultaneously related to two others that are not themselves also related. A simple example of this is a mode choice among private automobile, bus, and train: an auto and a bus are similar in that they share the same street network for travel, and the bus and train are similar in that they are both public transit, but the train and the auto are not related in either of these ways. The BL model is not constrained to such a hierarchical nesting of alternatives, but instead collapses to prevent any hierarchy. As a result, the relationships between alternatives are governed not by the depth, but instead by the breadth of their common nests.

The BL model, while difficult to estimate, also holds promise for simpler forecasting. This is because the conditional probabilities for each nest are binary (i.e., one for the best alternative and zero for all others), and determined exclusively by the best fractional alternative allocated to that nest. For certain types of choices, the "best" alternative can be calculated quickly. For example, in a route choice problem, the best choice in each nest can be found using a very efficient shortest path algorithm, and all of the remaining choices in the nest can be ignored. In the case of a BL model where $\alpha_{ir} = 0$ for all i, this can be especially advantageous, as it is no longer required to consider an inventory of all possible alternatives for forecasting. Instead, it is enough to identify the best-response (e.g. shortest path) alternative in each nesting block. Any other alternative is excluded from consideration. Even if $\alpha_{ir} > 0$, this can still be a useful tool, as the best-response alternatives might compete with a random sampling of other alternatives inside a regular MNL framework at the root level nest; the exact details of such a model are left for future work.

5 Application

Evaluating the block logit model in a theoretical sense is enlightening, but the true test of a model comes from how well it performs in modeling real world behavior. To provide an first look, the BL model was compared against a NL model using a work trip mode choice survey conducted by the Metropolitan Transportation Commission in San Francisco. In this dataset, six alternatives are considered: driving alone, sharing a ride with just one other person, sharing a ride with more than one other person, transit, biking, and walking. A relatively simple but reasonable utility function was constructed for use in both NL and BL models, consisting of cost, travel time, alternative specific income, and alternative specific constants.

ELM (ELM-Works, LLC., 2008) was used to estimate parameters for the various models. Various nesting structures were considered, including nesting together non-motorized modes, motorized modes, private auto modes, and just shared ride modes. For the NL model, only the shared ride nesting structure resulted in a logsum parameter within the utility maximizing acceptable range, between 0 and 1. The BL model is unable to reproduce a likelihood function that is similar to a NL model with an out-of-range logsum parameter, as the blocks are specifically constrained to prevent this. Because of this, the results for the corresponding BL models could not "match" the unacceptable NL results, and generally converged back to the MNL result.

One simple nesting structure, relating the shared ride 2 and shared ride 3+ alternatives, did generate an acceptable logsum parameter in the NL form. The results for this model, along with the corresponding MNL and BL models, are shown in Table 1. It can been easily seen that, in this particular case, the BL model results in a slightly better fit than the NL model, as represented by the superior value of the log likelihood.

As the NL and BL models have the same number of parameters, the increase in $\bar{\rho}^2$ of 2.35e-04 has a significance of less than 0.0318 (using the non-nested hypothesis test outlined in Ben-Akiva and Lerman (1985)), allowing the BL model to statistically reject the NL model with reasonably high confidence. However, from a practical standpoint, the BL model is not much different from the NL model. Neither model substantially changes the β parameter estimates from the MNL model form, and both induce similar mid-level correlations between the shared ride 2 and shared ride 3+ error terms.

During the estimation process, however, the NL formulation was superior. It found an optimum value for log likelihood easily, and converged quickly to the same optimum from multiple different starting points. The BL formulation, on the other hand, frequently became stuck in suboptimal local maxima, typically close to the MNL values. The (probable) global optimum was only found by estimating parameters on the model with a fixed μ value of 0.5, and then reducing it gradually to 0.005 and re-optimizing several times.

6 Discussion

This paper introduced the basic mathematical form of the block logit model, deriving it by placing restrictions on the network GEV model that mirror the restrictions used to generate the nested logit model. Some properties resulting model's log likelihood function were reviewed, and advantages and drawbacks considered. An anecdotal application of the BL model showed a statistical improvement over the NL model, but at significant cost in the parameter estimation process.

Table 1: Model Estimation Results

		Multinomial Logit			**Nested Logit**			**Block Logit**		
Log Likelihood Parameters		-3626.19			-3623.84			-3622.12		
Log Likelihood Null Model		-7309.6			-7309.6			-7309.6		
ρ^2		0.503914			0.504236			0.504471		
$\bar{\rho}^2$		0.502272			0.502457			0.502692		
Parameter		**Estimate**	**StdError**	**t-Stat**	**Estimate**	**StdError**	**t-Stat**	**Estimate**	**StdError**	**t-Stat**
Cost (cents)		-0.00492	0.000239	-20.6	-0.00481	0.000242	-19.9	-0.00482	0.000210	-22.9
Travel Time (minutes)		-0.0513	0.00310	-16.6	-0.0511	0.00307	-16.6	-0.0510	0.00247	-20.6
Alt. Specific Constants	Drive Alone	0	-	-	0	-	-	0	-	-
	Share Ride 2	-2.18	0.105	-20.8	-2.10	0.103	-20.4	-2.175	0.0967	-22.5
	Share Ride 3+	-3.73	0.178	-21.0	-3.17	0.225	-14.1	-3.31	0.108	-30.8
	Transit	-0.671	0.133	-5.06	-0.672	0.132	-5.09	-0.676	0.125	-5.39
	Bike	-2.38	0.305	-7.80	-2.37	0.304	-7.79	-2.37	0.302	-7.84
	Walk	-0.207	0.194	-1.07	-0.206	0.194	-1.06	-0.211	0.182	-1.15
Income ($000/yr)	Drive Alone	0	-	-	0	-	-	0	-	-
	Share Ride 2	-0.00217	0.00155	-1.40	-0.00185	0.00147	-1.26	-0.00172	0.00140	-1.23
	Share Ride 3+	0.000358	0.00254	0.141	-0.000588	0.00201	-0.293	-0.000896	0.00143	-0.626
	Transit	-0.00529	0.00183	-2.89	-0.00517	0.00182	-2.84	-0.00522	0.00182	-2.87
	Bike	-0.0128	0.00532	-2.41	-0.0128	0.00532	-2.40	-0.0128	0.00532	-2.40
	Walk	-0.00969	0.00303	-3.19	-0.00968	0.00303	-3.19	-0.00968	0.00303	-3.19
Shared Ride Logsum Parameter					0.656	0.107	-3.20[a]			
Shared Ride Allocation								0.338[b]	see note	-

[a] Calculated relative to a reference value of 1.

[b] ELM estimated allocation parameters in a log form, but the value is presented in the table in a manner consistent with the formulations described in this paper. The original estimated value for the allocation to the nest, holding the allocation to the root constant at zero, was -0.672, with a standard error of 0.296.

Overall, the block logit model shows little promise for general use in the discrete choice modeler's toolbox. It is difficult to work with, requires careful monitoring in parameter estimation, and has parameters that may generate results that do not have easily discerned effects on choice. The study of the block logit model does provide useful insights into the behavior of similar models, especially the CNL model, when their parameters approach extreme values, and may serve as a warning to users of those tools to be cautious in their application.

However, despite its obvious flaws and drawbacks, the BL model may be useful for particular specialized applications, where the ease of calculating nest probabilities can be leveraged to simplify calculations. If the difficult parameter estimation process can be overcome, the use of the block logit model for forecasting may make its implementation worthwhile. An empiracle evaluation of these potential benefits is left for future research.

References

Ben-Akiva, M. and B. François (1983) μ-homogeneous generalized extreme value model, Working paper.

Ben-Akiva, M. and S. R. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press Series in Transportation Studies, The MIT Press, Cambridge, MA.

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

Bresnahan, T., S. Stern and M. Trajtenberg (1997) Market segmentation and the sources of rents from innovation: personal computers in the late 1980s, *RAND Journal of Economics*, **28**, 17–44.

Chu, C. (1989) Paired combinatorial logit model for travel demand analysis, paper presented at *Fifth World Conference on Transportation Research*, vol. 4, 295–309, Ventura, CA.

Daly, A. and M. Bierlaire (2006) A general and operational representation of generalised extreme value models, *Transportation Research Part B: Methodological*, **40**, 285–305.

ELM-Works, LLC. (2008) ELM (Easy Logit Modeler), Version 1.0, http://www.elm-works.com.

McFadden, D. (1978) Modeling the choice of residential location, in A. Karlqvist (ed.) *Spatial Interaction Theory and Residential Location*, 75–96, North-Holland, Amsterdam.

Newman, J. P. (2008a) Normalization and disaggregation of networked generalized extreme value models, Ph.D. Thesis, Northwestern University, June 2008.

Newman, J. P. (2008b) Normalization of network generalized extreme value models, *Transportation Research Part B: Methodological*, **42** (10) 958–969.

Small, K. A. (1987) A discrete choice model for ordered alternatives, *Econometrica*, **55**, 409–424.

Vovsha, P. (1997) The cross-nested logit model: application to mode choice in the tel-aviv metropolitan area, *Transportation Research Record*, **1607**, 6–15.

Wen, C.-H. and F. S. Koppelman (2001) The generalized nested logit model, *Transportation Research Part B: Methodological*, **35** (7) 627–641.