NORTHWESTERN UNIVERSITY

Normalization and Disaggregation of Networked Generalized Extreme Value Models

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Civil and Environmental Engineering

By

Jeffrey P. Newman

EVANSTON, ILLINOIS

June 2008

Abstract

Normalization and Disaggregation of Networked Generalized Extreme Value Models

Jeffrey P. Newman

Generalized extreme value (GEV) models provide a convenient way to model choice behavior that is consistent with utility maximization theory, but the development of specific new models within the GEV family has been slow, due to the difficulty of ensuring new formulations comply with all the GEV rules. The network GEV structure (NetGEV) introduced by Daly and Bierlaire (2006) provides a tool to verify that proposed new models satisfy the GEV conditions, without the burden of complex analysis of the new model to ensure its properties. This dissertation further develops and expands the NetGEV tool. It describes several methodologies for applying constraints to correctly normalize the allocation parameters in such models, allowing parameter identification while ensuring that utilities are not biased due to the network structure. These methods vary depending on the structure of the underlying network.

Additionally, a modification of the allocation parameters is presented, which transforms them to create an alternative set of parameters that are unconstrained. This change also allows the inclusion of data within the allocation formulations, which creates a new heterogeneous network GEV (HeNGEV) model, with the opportunity for heterogeneous covariance structures, while maintaining the closed form probabilities common to GEV models. Including the heterogeneity

in the allocation structure, as opposed to the logsum parameters (as in Bhat, 1997a), allows variations in both the magnitude and structure of the covariance. This heterogeneity is useful in sub-market analysis, where small differences in the competitive dynamic between alternatives in a segment of the population may drive large changes for revenue management systems or environmental justice evaluations. Various derivatives and elasticities of the HeNGEV model are derived, utilizing the network structure underlying the model to simplify the formulations.

The performance of the HeNGEV model is compared against a homogeneous NetGEV model, using two different synthetic data sets. The first data set is designed to maximize the effect of the heterogeneous error covariance, while the second reduces the effect to a more subtle level. In each case, the HeNGEV model performs better than the NetGEV model, recovering parameters that are closer to their (known) true values, and with improvements in log likelihoods well above a statistically significant threshold.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

The modeling of choice processes is an important part of transportation demand forecasting, as transportation system demand is derived from individual choices made by travelers. Travel demand forecasting underlies every step of transportation planning, as the evaluation of policy and planning initiatives requires the anticipation of how travelers will react to changes in the transportation system.

Early demand forecasting models made assumptions about behavior that were unrealistic, but allowed the models to be calculated relatively simply, using technology available at the time. In particular, one of the earliest disaggregate choice models was the multinomial logit model, which exhibited the property of independence from irrelevant alternatives (IIA). The resulting choice probabilities, and in particular trade-offs between alternatives, were obviously not consistent with realistic choices. Specifically, the probabilities of all other alternatives changed proportionally in response to the inclusion, exclusion or change in any alternative. Yet the MNL model was (and in some applications, continues to be) a popular tool, because of the ease of estimation and application.

1.1. Generalized Extreme Value Models

The problems of the MNL model were well known early in its application, but the few quantitative alternatives were extremely complicated and extraordinarily difficult to estimate or to be used for prediction. This impasse existed until the development of the generalized extreme value

(GEV) structure for discrete choice models proposed by McFadden (1978). This framework established a set of rules that defined a family of models, which included the MNL, but also many other forms that were not impeded by the IIA property, by incorporating non-independent covariance structures for the error terms.

Since the development of the GEV structure, substantial efforts have been put forth to find new forms of GEV model, exhibiting more varied covariance structures. Progress was initially slow, and for some time modelers were limited to the initial multinomial logit and nested logit models, which both pre-dated the more general GEV formulation. Later, Chu (1989) added the paired combinatorial logit, Vovsha (1997) the cross-nested logit, Small (1987) the ordered GEV, and Bresnahan et al. (1997) the product differentiation model. Ultimately, Wen and Koppelman (2001) proposed the generalized nested logit (GNL) model, which is a more general form which encompasses all previous such models, with the exception of the multi-level nested logit model. The GNL, unlike the nested logit model, is limited to only a single level of nests, and does not allow hierarchical nesting.

The pace of discovery of new GEV models was impeded primarily because each new model needed to be carefully constructed to ensure compliance with the GEV prerequisites, but also because more complex models generally required substantial computational effort. Technological advancements in computing power and data storage have thus made it possible to estimate ever more detailed and complex models. For example, Coldren and Koppelman (2005) introduced a three level weighted nested logit model, as well as a nested weighted nested logit model. Each of these was essentially a multi-level expansion of the GNL model, but they were specific cases of models, which lacked generalizability to more abstract multi-level forms.

While it is relatively easy to envisage a particular correlation structure among alternatives, it is sometimes hard to translate that structure into a viable GEV formulation. In particular, the requirement that the generating function has alternating sign partial derivatives is not trivial to check for most possible generating functions. In light of this, several authors have examined methods of stitching together separate GEV models to create new ones, allowing modelers to use these tools to develop models with new utility correlation structures on the fly. Such tools would obviate the need to carefully analyze each new model structure to ensure it is compliant with the GEV formulation. The finite mixture model developed by Swait (2003) merged together several models to create a new GEV model. Daly and Bierlaire (2006), hereafter referred to as D&B, separately expanded this idea to include not only connecting separate models, but also creating a new, more flexible structure for making new GEV model forms by connecting partial GEV models using a network (NetGEV). D&B's network formulation is one of the most general of GEV model structures discovered so far, as all other models mentioned above are specific forms of the NetGEV model. Refining and expanding this model is the focus of this dissertation.

1.2. Advanced GEV Models

There are, however, certain GEV constructs that are not specific forms of the NetGEV model. Most notable is the mixed logit form, which is very flexible, and can approximate (to any arbitrarily close degree of precision) any possible GEV form (Revelt and Train, 1998; McFadden and Train, 2000; Train, 1998, 2003). The primary drawback of mixed logit models is that they do not have a closed form probability expression, as the various models mentioned above do. The lack of a closed form means that finding probabilities requires calculating integrals, a task usually handled through simulation. While the conceptual form of the mixed logit model has been known for many years, it has only been in the past decade or so that computational power has been available for even the most rudimentary mixed logit models to be estimated (Train, 2003).

Mixed logit models have been used principally to allow heterogeneity of the utility functions across individuals. This heterogeneity can be included in the systematic portion of utility (in a random coefficients formulation), or in the random portion of utility (in an error components formulation). While these two forms are mathematically the same within the mixed logit framework (Train, 2003, p. 144), their interpretation can be different, depending on what variables are included in different terms of the utility.

Latent class models can incorporate heterogeneity for systematic utility into a closed form model. These models differ from what is traditionally considered as a mixed logit model by having discrete distributions on the random parameters, instead of continuous distributions. Such models have been used extensively in marketing (e.g. Kamakura and Russell, 1989; Chintagunta et al., 1991; Swait, 1994; Schrevens et al., 2005), as well as transportation choices such as wilderness recreation locations (Boxall and Adamowicz, 2002) and intercity mode choice (Bhat, 1997b).

While substantial work has been done to bring heterogeneity into the systematic portion of utility, relatively little has been done to incorporate heterogeneity into the random portion of utility. One of the reasons that modelers adopt a mixed logit formulation is the ease of incorporating differences in decision makers into the error component of utility. Bhat (1997a) showed that it is possible to incorporate such differences in a closed form GEV model, which he eloquently named "nested logit model which accommodates covariance heterogeneity", by sub-parameterizing the logsum parameters of a nested logit model. This model allowed differing *amounts* of covariance across people, but required a single covariance nesting *structure* throughout the population; so that there could be more or less covariance for different people, but not a different nesting structure entirely.

1.3. Motivation for Further Improvements

The most flexible discrete choice models available today are mixed logit models, which can achieve any desired correlation structure among the random utilities of choices. However, using such models requires simulation, an expensive computational process. The costs of simulation, while continually shrinking with advances in computer capabilities, are generally over burdensome for more complex models, such as activity-based models (Planning Section MTC, 2005) or integrated land use and transportation models (Waddell et al., 2007). Even larger models, such as the travel models embedded in *Epicast*, a national-scale epidemiological simulation (Germann et al., 2006), simply cannot handle the storage of individual-specific parameters that a mixed logit model requires, even on the most advanced super computers in existence today. Such applications call for closed form models, which can generate choice probabilities without the costs of simulation.

Nevertheless, it is desirable to incorporate into closed form models as many of the features of the most flexible mixed logit models as possible. In particular, a heterogeneous error covariance structure, which heretofore has been only available with a mixed logit model, would allow a wider array of inter-alternative interactions, and potentially better fitting models.

The information available in observed data is often limited, especially with shrinking budgets for data collection. When data collection is relatively complete, the various ways in which alternatives are similar or different are captured in the systematic portion of utility, with the random portion of utility left to represent only uncorrelated "white noise" in the choice process (Train, 2003, p. 39). It is when the data is incomplete that more sophisticated models can fill in some of the gaps. The development of choice models with more flexible correlation structures for the random portion of utility allows for the capture of more information out of the unobserved portion of utility.

1.4. Contributions

The focus of this dissertation is the development of a closed form GEV model which exhibits covariance heterogeneity across decision makers, both in the quantity of covariance and in the structure of that covariance across alternatives. Using the NetGEV model introduced by D&B as a jumping off point, this work provides several specific contributions:

- A demonstration is provided that shows that the correct normalization of allocation parameters in a NetGEV model, in the absence of a complete set of alternative specific constants, depends on the structure of the network.
- For two particular network topologies of NetGEV models, non-biasing normalizations for the allocation parameters are introduced. For networks not consistent with either of these topologies and lacking a full set of alternative specific constants, a remedial transformation is provided that still allows non-biasing normalization, although through the use of complicated non-linear parametric constraints.
- An alternative functional form for the NetGEV model is proposed, that replaces constrained allocation parameters with unconstrained parameters, which are simpler to estimate using common maximum likelihood methods.
- Lastly, a new closed form heterogeneous covariance network GEV model (HeNGEV) is presented, which incorporates decision maker characteristics into node allocations, allowing such characteristics to drive not only the systematic utility of the alternatives, but also the structure and magnitude of the error correlations.

1.5. Benefits

Improved behavioral realism. A major benefit of any relaxation in the assumptions required by a choice model is an improvement in the potential behavioral realism of the model. The specification of random utility maximization models requires some assumptions about the distribution of utility across alternatives. The basic MNL model makes some fairly restrictive assumptions about that distribution, namely that the utilities for the alternatives are distributed identically and independently with a Gumbel distribution. The nested logit and generalized nested logit models each relax this assumption, allowing (respectively) hierarchical or overlapping sets of alternatives to be non-independent, with homogeneous correlation across the population. The NetGEV model allows correlation across both hierarchical *and* overlapping sets of alternatives simultaneously. The introduction of the HeNGEV model relaxes the homogeneity of correlation assumption. Each relaxed assumption allows the models potential to more closely match the underlying choice process.

Overcoming data deficiencies. Obviously, it is desirable to include in a random utility model as many of the variables that determine utility as possible. In the ideal case, all such variables will be included, and the remaining error would have a tiny variance, which would be uncorrelated across alternatives. Such a model could be fully captured with a regular multinomial logit structure. However, in practice we nearly always lack all the relevant variables that determine utility. Some variables are unknown or difficult to quantify, and others are too expensive and time consuming to collect for each sampled decision maker. The error term thus must encompass the effect of a wide variety of unobserved attributes of the alternatives.

The covariance structure of the error terms in a random utility model represents the similarities across alternatives in unobserved variables. Early members of the GEV family of models allowed this covariance structure to differ from zero, but still imposed significant constraints on the form and quantity of covariance, which translates to an implied constraint on the nature of the similarities in the unobserved attributes of the alternatives. Relaxing that constraint will allow models to more closely approximate the real underlying choice mechanism. The NetGEV model is able to represent a wider array of possible correlation structures than the most general previously known GEV models (for details, see Section 2.5 on page 28). This allows the incorporation into the model of more of the effect of these unobserved attributes of the alternatives, which can mitigate the loss of modeling predictive power that comes from not including those unobserved attributes directly.

Enhanced market segmentation. The heterogeneous covariance introduced in this dissertation is an extremely valuable tool for understanding the competitive dynamics among alternatives in segments of the population. In traditional GEV models, various segments of the population can reveal different preferences for alternatives though parameters within the systemic portion of utility. This individual response function is the basic nature of disaggregate models. However, the random portion of utility in these models is homogeneous, so that it cannot reflect differences among decision makers in the population. Allowing the error terms to have heterogeneous covariance is a natural extension of disaggregate choice models.

This extension can be important in evaluating the effects of policy changes on particular subsegments of the population. It is well known that the change from a fixed (independent) covariance structure to a parameterized structure, as when moving from a multinomial logit to a nested logit, can change the resulting probabilities generated by the model across the entire population. Incorporating a heterogeneous covariance is the functional equivalent of changing from the fixed covariance to a parameterized covariance across a continuum of submodels spanning the values of the variables that are included in the parameterization.

Changing from a homogeneous covariance to a heterogeneous covariance would not typically be expected to reveal large changes in probability across the entire population, as the likelihood maximizing parameters for the homogeneous model would provide a balance of the covariances, even when the covariances differ across the population. However, as is demonstrated in this dissertation, the heterogeneous covariance could result in very different probabilities in subsections of the population, if the underlying covariance is indeed heterogeneous.

This improvement in predictive ability for market segments will be of particular interest to modelers that are focusing only on subsections of the population. Within revenue management programs, which specialize in extracting rents from market niches, the heterogeneous covariance structure will be particularly appealing. Applying a heterogeneous model to itinerary choice could reveal choice patterns and trade-offs among certain (especially high income) travelers that are not reflected in more general choice models.

Heterogeneous covariance could also be relevant in environmental justice issues. The U.S. federal government requires evaluation of proposed projects regarding their impact on disadvantaged groups. If individuals in these groups tend to see particular sets of alternatives as more, or less, similar, when compared against the rest of the population, then the impact of projects on such groups may be misrepresented by a homogeneous analysis. For example, if lower income people tend to see bus and rail transit as very similar, while higher income people do not, then a project to enhance rail service might have benefits that accrue more to higher income people than would be predicted by a homogeneous model.

Closed form model. The various benefits described above could be achieved through the implementation of mixed logit models. However, the use of mixed logit models requires multidimensional integration through simulation, which can be a computationally expensive process. The closed form of the HeNGEV model requires substantially less computation.

Graphical representation. Additionally, the specification of a mixed logit model that captures a complex covariance structure requires careful consideration of the functional form of parameters and errors. The HeNGEV structure, on the other hand, inherently provides a graphical representation of the covariance structure, with well defined relationships. While not a straightforward and simple as the nested logit form, it is still easier to grasp for a modeler than a completely amorphous form.

1.6. Outline

In the next chapter, the details of GEV models and the NetGEV structure proposed by D&B, as well as a convenient mathematical reformulation of this structure, are reviewed. Topological properties of NetGEV structures, and some possible normalizations of parameters in those structures are then presented in Chapter 3. A useful transformation of the allocation parameters, which results in a novel heterogeneous covariance network GEV model (HeNGEV) is introduced in Chapter 4. This new HeNGEV model eases estimation by removing parametric constraints, and creates the possibility of the inclusion of disaggregate data into the utility allocations, allowing the structure of the network, and thus the nature of the underlying correlation structure of the random utility values, to vary within the population. The various derivatives and the elasticity of probability with respect to observed attributes of the choices and the decision maker for the HeNGEV model are examined in Chapter 5. Finally, an operational application of the HeNGEV model is demonstrated, using synthetically generated airline itinerary choice data in Chapter 6. The final chapter provides a summary of the presented innovations, and some concluding remarks.

CHAPTER 2

Modeling Structures

2.1. Random Utility Maximization

The most common method for modeling such choices is through random utility maximization, a methodology formalized by Manski (1977). This type of model is based on utility maximization theory, where an individual is surmised to choose from any set of alternatives that which would provide him with the maximum benefit, or utility. The choices actually made by an individual are presumed to be deterministic: each person knows the potential utility of each alternative, and chooses the best one. The modeler, however, cannot directly observe utility, but instead can only observe some of the attributes of the alternatives that contribute to, but do not completely determine, utility. Those attributes provide a modeler some inputs with which to determine a probability density function representing the possible values of utility for the alternatives in an individual choice. The resulting model yields choice probabilities such that

$$P_{ti} = \text{Pr}(U_{ti} \geq U_{tj}, \forall j \in \mathcal{C}), \tag{2.1}$$

with P_{ti} as the probability that decision maker t chooses alternative i from choice set $\mathcal{C}$. Thus

$$P_{ti} = \int_{U_{ti}=-\infty}^{+\infty} \int_{U_{tj}=-\infty}^{U_{ti}} f\left(U_t\right) dU_t \tag{2.2}$$

The details of the model then depend on the functional form of U. Generally, the utility of an alternative in a model is separated into two components: a systematic, deterministic portion of utility that is derived from the observed attributes of the alternatives and the decision maker (V_i);

and a random component ε_i that is not attributable to observed attributes. Thus (2.1) becomes

$$P_{ti} = \Pr(V_{ti} + \varepsilon_{ti} \geq V_{tj} + \varepsilon_{tj}, \forall j \in \mathcal{C}).$$

2.2. Generalized Extreme Value Models

McFadden (1978) found that if the probability density function of the random portion of utility is assumed to be a multivariate generalized extreme value (GEV) distribution adhering to certain conditions, then the probability function can be reduced to a closed form expression. In addition, McFadden demonstrated that such distributions could be generated by other functions, when those generating functions $G(y)$ adhered to a few simple rules, to wit:

- $G(y) \geqslant 0, \ \forall y \in \mathbb{R}^J_+$,
- G is homogeneous of degree $\mu > 0$,[1]
- $lim_{y_i \to +\infty} G(y) = +\infty$, and
- the mixed partial derivatives of G with respect to elements of y exist, are continuous, and alternate in sign, with non-negative odd-order derivatives, and non-positive even-order derivatives.

When y in such a generating function is replaced with exp(V), then the resulting choice model has a closed form probability expression, and is consistent with random utility maximization theory. Different generating functions will result in different probability density functions within the generalized extreme value family. The primary benefit of varying the generating function is that different generating functions will result in multivariate density functions with different attributes, in particular with different covariance matrices. The ability to incorporate covariance

[1]McFadden (1978) originally required that G had to be homogeneous of degree 1, but this condition was relaxed by Ben-Akiva and François (1983), such that G needs only be homogeneous of any positive degree.

between the random portion of utility allows the modeler to partially account for relationships between alternatives that are not expressed in the observed characteristics of those alternatives.

Since the development of the generalized extreme value (GEV) structure for discrete choice models by McFadden (1978), substantial efforts have been put forth to find new forms of GEV model, exhibiting more varied covariance structures. Progress was initially slow, and for some time modelers were limited to the initial multinomial logit (with $G(y) = \sum_i y_i$) and nested logit models (in their simplest form, $G(y) = \sum_n [(\sum_{i \in n} y^{1/\mu_n})^{\mu_n}]$), which both pre-dated the more general GEV formulation. Later, Chu (1989) added the paired combinatorial logit, Vovsha (1997) the cross-nested logit, Small (1987) the ordered GEV, and Bresnahan et al. (1997) the product differentiation model. Ultimately, Wen and Koppelman (2001) proposed the generalized nested logit (GNL) model, which is a more general form which encompasses all previous such models, with the exception of the multi-level nested logit model. The GNL is derived from the generating function

$$G(y) = \sum_n \left(\sum_{i \in n} \left(\alpha_{ni} y_i\right)^{1/\mu_n} \right)^{\mu_n},$$

with the conditions that $\alpha_{ni} \geq 0$, and $\sum_n \alpha_{ni} = 1,\ \forall i$. The GNL, unlike the nested logit model, is limited to only a single level of nests, and does not allow hierarchical nesting. This process of discovery of new GEV models was initially slow, primarily because each new model needed to be carefully constructed to ensure compliance with the GEV prerequisites, but also because more complex models generally required substantial computational effort. Technological advancements in computing power and data storage have thus made it possible to estimate ever more detailed and complex models. For example, Coldren and Koppelman (2005) introduced a three level weighted nested logit model, as well as a nested weighted nested logit model.

2.3. D&B's Network GEV Formulation

Developing new members of the GEV family has been a difficult process, because the requirement that the generating function has alternating sign partial derivatives is not trivial to check for most possible generating functions. However, D&B propose a method to build new GEV models by combining other GEV models, based on the following functional form:

$$G^i : \mathbb{R}_+^{J_i} \to \mathbb{R}_+ : G^i(y) = \sum_{j \in i^{\downarrow}} \left[a_{ij} \left(G^j(y) \right)^{\mu_i/\mu_j} \right], \tag{2.3}$$

where i is a node in a finite, directed, connected, circuit-free graph, $i^{\downarrow}$ is the set of successor nodes to i, a is an allocation parameter associated with each edge in the network, μ is a scaling parameter associated with each node in the network, and J_i is the number of alternatives of the choice at i. The network is structured in such a way that there is a set of nodes with no successors, which correspond to the alternatives, and a single root node with no predecessors, that will directly or indirectly incorporate the G functions of all other nodes in the network, and thus the G function of that node represents the complete choice model. At the nodes corresponding to the choice alternatives (which have no successor nodes), equation (2.3) is replaced instead with

$$G^i : \mathbb{R}_+^{\dim_{\mathbb{R}}(y)} \to \mathbb{R}_+ : G^i(y) = y_i^{\mu_i}. \tag{2.4}$$

D&B show that if the allocation parameter a associated with each edge in the network is greater than zero, and if the scaling parameter μ associated with each node i is smaller than or equal to the scaling parameter of all successor nodes, then the G function at the root node will be a valid GEV function. While the relative scaling of this formulation (logsum parameters that grow when moving from the model root toward the alternatives, instead of shrinking) may seem "backwards" compared to the more common ordering of logsum parameters in a nested logit model (as in Train, 2003), this is merely a technicality of the particular formulation chosen by D&B, and is not mathematically limiting. Inverting the power of G^j in (2.3) and simultaneously inverting μ for every node does not change the net power of G^j; the resulting model is the same, but with the restriction that μ decreases when moving from the root node to the elemental alternatives, instead of increasing. The important criteria is that magnitude of the power scaling of the terms of the summation in (2.3) must not shrink when moving from the root node towards the elemental alternative nodes. The magnitude of the power of the terms contained in a G is inversely related to the variance of the alternatives (or portions of alternatives) contained in that G. For any set S of alternatives (or portions of alternatives), any subset of S must have equal or smaller variance, hence the required ordering of logsum parameter values.

2.4. An Alternative Mathematical Formulation

It will be useful to adjust this network structure, to aid in the simplification of the necessary constraints for normalization. The adjustments are similar to those made on the cross-nested logit model by Abbé et al. (2007).

Let $\mathcal{G}(\mathcal{Z}, \mathcal{E})$ be a finite connected directed circuit-free graph, in which only one node R in $\mathcal{Z}$ has no predecessor. Each edge in the graph is associated with a positive parameter $a_{ij} > 0$. We consider the subsets of nodes: $\mathcal{C}$ is the set of nodes with no successor, and $\mathcal{N}$ is the set of nodes with at least one successor, i.e. the complement of $\mathcal{C}$ in $\mathcal{Z}$. Since $\mathcal{G}$ is connected, R is a member of $\mathcal{N}$. We associate with each node i in $\mathcal{N}$ a parameter $\mu_i > 0$. Further, if a directed edge connects from any node i to any other node j (i.e. i is a direct predecessor of j), then $\mu_i \geq \mu_j$. If $i \in \mathcal{C}$ then

$$G^i : \mathbb{R}_+^{\dim_{\mathbb{R}}(y)} \to \mathbb{R}_+ : G^i(y) = y_i,$$

otherwise $i \in \mathcal{N}$ and

$$G^i : \mathbb{R}_+^{\dim_{\mathbb{R}}(y)} \to \mathbb{R}_+ : G^i(y) = \left(\sum_{j \in i^{\downarrow}} \left[\left(a_{ij} G^j(y) \right)^{1/\mu_i} \right] \right)^{\mu_i}, \tag{2.5}$$

where $i^{\downarrow}$ is the set of successor nodes of i. Let such a network be called a *scale-normalized GEV network*.

Theorem 2.1. *The function G^i associated with any node i of a scale-normalized GEV network is a 1-GEV function.*[2]

Proof. By D&B's Theorem 4, G^i is a 1-GEV function if

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

is a $1/\mu_i$-GEV function. By D&B's Theorem 1, H^i is a $1/\mu_i$-GEV function if all the terms in the summation are $1/\mu_i$-GEV functions. For each node $j \in i^{\downarrow}$, either $j \in \mathcal{C}$ or $j \in \mathcal{N}$. If $j \in \mathcal{C}$, expanding that term of H^i yields $(a_{ij} y_j)^{1/\mu_i}$, which is clearly a $1/\mu_i$-GEV function. If $j \in \mathcal{N}$, expanding that term of H^i yields $a_{ij}^{1/\mu_i} \left(H^j(y) \right)^{\mu_j/\mu_i}$, which by D&B's Theorem 4 is a $1/\mu_i$-GEV function if $H^j(j)$ is a $1/\mu_j$-GEV function. By induction, all nodes i in the network have G^i as a 1-GEV function, and all nodes $j \in \mathcal{N}$ have an associated H^j, which is a $1/\mu_j$-GEV function. □

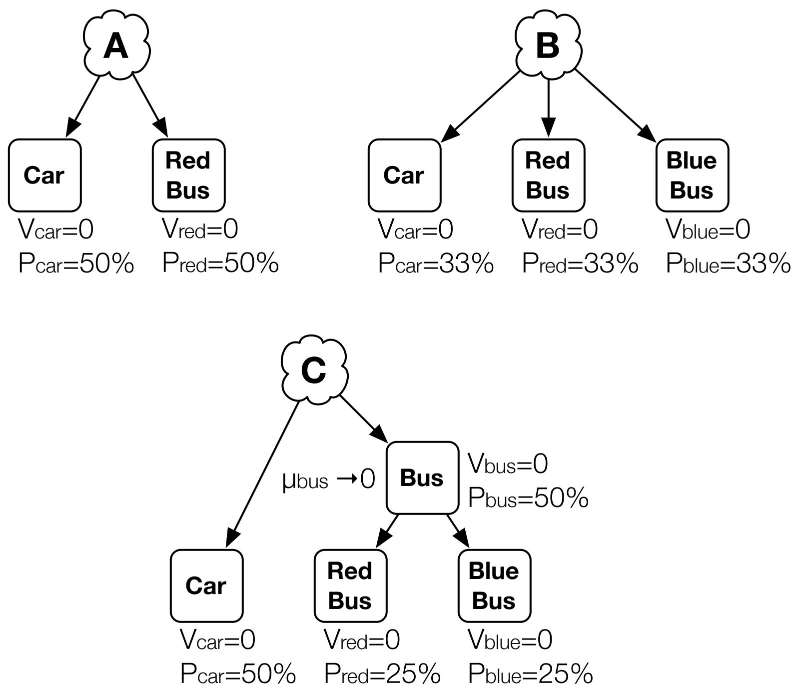
[2] Strictly interpreted, G^i for any node that does not contain the entire set of alternative nodes $\mathcal{C}$ is not exactly a GEV generating function, but merely an asymptotic approximation of a GEV function. This is a result of the fact that for any elemental alternative node j not in the successor set of i, $\lim_{y_j \to +\infty} G^i(y) \neq +\infty$, violating one of the formal requirements of a GEV function. The result of this violation is that the function generated by G^i is not strictly a CDF, as what should be the marginal cumulative density function for the associated elemental alternative j is $F_j = 1$. However, F_j is an asymptotic approximation the univariate extreme value distribution $F_j = \exp\left[-\exp\left[\mu - \varepsilon_j\right]\right]$ as $\mu \to -\infty$, and thus G^i is an asymptotic approximation of a GEV generating function. Since the only model typically used is the model associated with the root node at the beginning of the network, and since this node by construction will have all the elemental alternative nodes as eventual successor nodes and thus will have a fully compliant GEV generating function, this distinction should be irrelevant in practice, but it is included here for completeness.

The resulting model that is derived from this network is mathematically equivalent to the model presented by D&B, as can be seen in the similarity between Theorem 2.1 and D&B's Theorems 7 and 10. There are, however, a few useful differences:

- The G^i function at every node is a 1-GEV function. Since the scale of G^i here is the same at every node of the network, direct "apples to apples" comparisons of G^i values for any two nodes can be made. While this has no impact on the resulting choice probabilities, it can make the interpretation of complex networks simpler, by allowing the direct identification of preferred network paths through simple numerical comparisons of utilities.
- The μ parameters attached to nodes in $\mathcal{C}$ are dropped from the model, as they are not identifiable, as discussed in the next section. This is *not* the same as normalizing μ from the choices end of the network, but rather because these parameters are not mathematically relevant.
- The allocation parameters, which D&B denote as α, have been denoted here as a, and they have been applied to G before the μ power transformation, rather than after. This change simplifies the normalization restrictions described in the next section, as a will be a function of the ultimate parameters α, and will have different functional forms depending on the structure of the network.

The relationship between G^i and V_i, the systematic utility of the alternative, is simple when node i is an elemental alternative ($G^i = \exp[V_i]$). It is useful to conceptualize a similar relationship between G^n and V_n for nesting nodes, even though those nodes do not have a direct systematic utility per se. V_n for nesting nodes is still a relevant measure of utility. In the nested logit model, V_n is the scale adjusted logsum value for the nest. It retains a similar function in the NetGEV structure.

Figure 2.1. One Bus, Two Bus, Red Bus, Blue Bus



2.5. Advantage of NetGEV over GNL

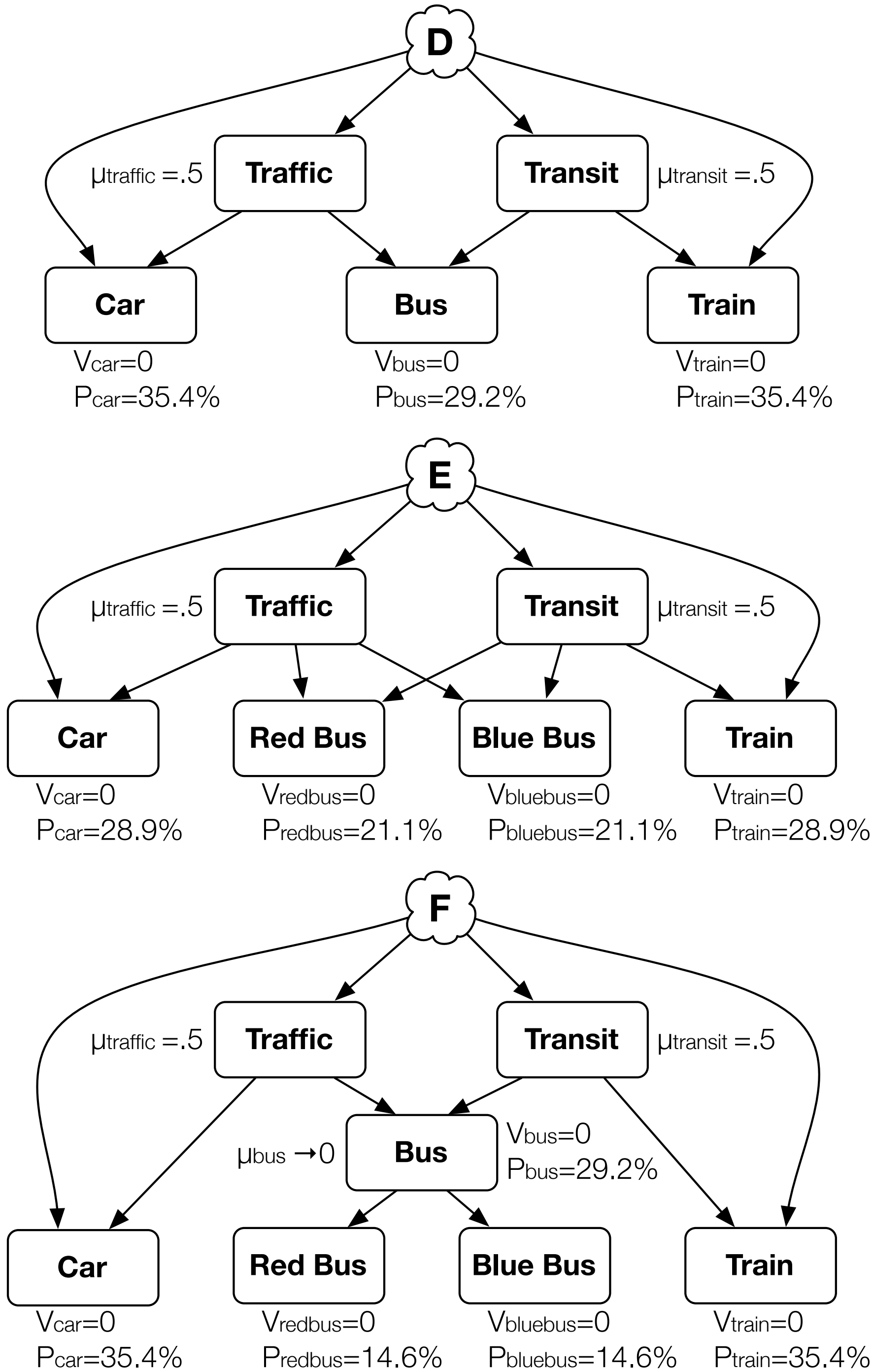
The NetGEV model is more flexible than the GNL model, and is able to represent a greater range of possible correlation structures between alternatives. In particular, the hierarchical nesting structure allows strongly correlated alternatives to still be loosely correlated with other alternatives. Wen and Koppelman (2001) begin to explore the differences between the GNL and the hierarchical form as expressed in the NL model, but they conclude that the GNL can generally approximate an NL model. The NetGEV model, on the other hand, makes these differences potentially more relevant.

For example, consider the famous red bus/blue bus problem. In the traditional scenario, a decision maker is initially faced with a choice between travelling in a car or in a red bus, as in the A model in Figure 2.1.In the simplest case, these alternatives are considered equally appealing, and each has a 50% probability of being chosen. When a new blue bus alternative is introduced, which is identical in every way to the red bus, we would expect the bus riders to split across the buses, but car drivers would not move over to a bus alternative. In the MNL model, however, this does not happen. Instead, as in the B model in Figure 2.1, the buses draw extra probability compared to the original case. The introduction of the nested logit model, as in the C model, allows the error terms for the bus alternatives to be perfectly correlated, and we achieve the expected result.

However, in a revised scenario, the original case is not binary choice, but instead it is a three-way choice, between a car, a bus, and a train. Further, we can construct the initial model as a GNL model (shown in the D model in Figure 2.2), so that the car and bus alternatives are partially nested together (both get stuck in traffic), and the bus and train alternatives are also partially nested together (both are mass transit). In this model, the utility of the bus tends to fall between car and train, so that its probability is slightly reduced relative to the others (a decision maker chooses the alternative with maximum utility, and the bus tends to be most likely to come out as #2 in this scenario). Again, the blue bus is introduced into the market, identical to the red bus. If the blue bus is inserted into the GNL model with the same nesting setup as the existing red bus, as in the E model in Figure 2.2, the probabilities of the car and train alternatives are adversely affected. A new "bus" nest could be introduced to induce the required perfect correlation between the error terms of the buses, but under the constraints of the GNL model, the allocations of the buses to the traffic and transit nests would need to be reduced (to zero), eliminating the correlation between the buses and the other alternatives.

The NetGEV model removes that constraint of the GNL model, and allows hierarchical nesting, as in a normal NL model. Thus, the nesting structure in the F model of Figure 2.2 can be created, linking together the buses before allocating them to traffic and transit nests. The probabilities for car and train can be preserved, with the red and blue buses splitting the bus market only.

Figure 2.2. The Blue Bus Strikes Again



CHAPTER 3

Normalization of Parameters

The NetGEV model as formulated is over-specified, so that is not possible to identify a unique likelihood-maximizing set of parameters. The over-specification is similar to that observed in attempts to maximize $f(x,y,z) = -(x+y)^2 + (z/z)$. This problem cannot be solved to an identifiable unique solution; any value for any individual parameter can be incorporated into a maximizing solution. Some parameters are unidentified as a set (as are x and y), and can only be identified if one of the set is fixed at some externally determined value (e.g. setting $y = 1$) or if some externally determined relationship is applied (e.g. setting $x = y$). Other parameters are intrinsically unidentified (in this example, z), and cannot be identified at all.

Mathematically, this is expressed in the derivatives of f with respect to its parameters. The first derivative of f with respect to an intrinsically unidentified parameter is globally zero. Parameters unidentified in sets can individually have calculable first partial derivatives, but the Hessian matrix of second derivatives is singular along the ridge of solutions.

In the D&B formulation, the μ parameters for the nodes associated with the elemental alternatives are intrinsically unidentified. For G^i of a node i in $\mathcal{N}$, the additive term relating to any node j in $\mathcal{C}$ is $a_{ij}G^j(y)^{\mu_i/\mu_j} = a_{ij}(y^{\mu_j})^{\mu_i/\mu_j} = a_{ij}y^{\mu_i}$, which drops μ_j. This can be verified by examining the derivatives of the D&B formulation. They provide the derivative of $G^i(y)$ with respect to μ_k when k is a successor node of i:

$$\frac{\partial G^i\left(y\right)}{\partial \mu_k} = a_{ik}\left(G^k(y)\right)^{(\mu_i/\mu_k)-1}\frac{\mu_i}{\mu_k}\left(\frac{\partial G^k(y)}{\partial \mu_k} - \frac{1}{\mu_k}G^k(y)\log\left[G^k(y)\right]\right). \tag{3.1}$$

When k is associated with an elemental alternative, then $\partial G^k(y)/\partial \mu_k = y_k^{\mu_k} \log[y_k]$. Inserting this into (3.1) will cause the term in parenthesis to collapse to zero, which will propagate through the rest of the network, making the derivative of this parameter globally zero.

3.1. Topological Reductions

The topographical structure of the GEV network can create further over-specification, by including extraneous nodes and edges that do not add useful information or interactions to the choice model. Fortunately, these extraneous pieces can be removed from the network without changing the underlying choice model.

3.1.1. Degenerate Nodes

A degenerate node is a node in the network that has exactly one successor. The G function for a degenerate node d collapses to a single term:

$$G^d(y) = \left(\sum_{j \in d\downarrow} \left[\left(a_{dj} G^j(y) \right)^{1/\mu_d} \right] \right)^{\mu_d} = a_{dj} G^j(y). \tag{3.2}$$

μ_d drops out of the equation, and has no effect on G^d, and thus no effect on any other G in the network, including G^R. Since μ_d disappears from the calculation, it is intrinsically unidentified. Degenerate nodes can, however, be removed from the network.

Theorem 3.1. (Degenerate Node Collapse) *Any degenerate node can be removed from the network, along with the edge e connecting from it to its successor, by redirecting all incoming edges from the degenerate node directly to the degenerate node's successor, and setting the allocation parameter on each such redirected edge equal to the product of the edge's original allocation and the allocation on e. This change will not affect the choice probabilities of the resulting model.*

PROOF. In the original network, if $d \in i^{\downarrow}$ then

$$G^i = \left(\left\{ a_{id} G^d \right\}^{1/\mu_i} + \sum_{m \in \{i^{\downarrow} \setminus d\}} \left[(a_{im} G^m)^{1/\mu_i} \right] \right)^{\mu_i}.$$

Replacing G^d with the rightmost side of (3.2) changes the term in curly braces to $a_{id} a_{dj} G^j$, resulting in an equation which exactly matches the correct form needed for the revised network, with $a_{ij} = a_{id} a_{dj}$. □

While certain non-normalized nested logit models may require degenerate nodes to correctly normalize the model (Koppelman and Wen, 1998), the NetGEV model does not require such nodes, and they can always be safely removed from the network. If they are not removed, it will be necessary to externally identify the value of any degenerate node's logsum parameter.

3.1.2. Vestigial Nodes

A vestigial node is a node which has no successors, but is not associated with an elemental alternative. While such nodes generally would not be expected in any practical application, the definition of a GEV network does not technically preclude their existence. The G function for such a node would always equal zero, as the set of successor nodes in the summation term of (2.5) is empty. The removal of such nodes from the network would obviously not affect the resulting choice probabilities. As with degenerate nodes, if they are not removed, it will be necessary to externally identify the value of their logsum parameters.

3.1.3. Duplicate Edges

Duplicate edges also add complexity to the network without providing any useful properties. A duplicate edge is any edge in the network that shares the same pair of ends as another edge. As the network is defined to be circuit free, all duplicate edges will always be oriented in the same direction. The allocation parameters on any set of duplicate edges are jointly unidentified, but the extra edges can be removed without altering the underlying choice model.

Theorem 3.2. (Duplicate Edge Removal) *For any set of D duplicate edges $e_1, e_2, ..., e_D$ in the network, which connect from node i to node k, $e_1, e_2, ..., e_D$ can be removed and replaced with a single edge $\hat{e}$, with the allocation parameter on $\hat{e}$ equal to $\left(\sum_{d=1}^{D} a_{e_d}^{1/\mu_i}\right)^{\mu_i}$. This change will not affect the choice probabilities of the resulting model.*

PROOF. In order to maintain the model probabilities, G^i must maintain the same value before and after the edge removal. So,

$$\left(\sum_{d=1}^{D}\left[\left(a_{e_d}G^k\right)^{1/\mu_i}\right] + \sum_{j\in\{i^{\downarrow}\setminus k\}}\left[\left(a_{ij}G^j\right)^{1/\mu_i}\right]\right)^{\mu_i} = \left(\left(a_{\hat{e}}G^k\right)^{1/\mu_i} + \sum_{j\in\{i^{\downarrow}\setminus k\}}\left[\left(a_{ij}G^j\right)^{1/\mu_i}\right]\right)^{\mu_i},$$

thus

$$\sum_{d=1}^{D}\left[a_{e_d}^{1/\mu_i}\right] = a_{\hat{e}}^{1/\mu_i},$$

and the result follows directly. □

When a GEV network has been stripped of degenerate and vestigial nodes, and duplicate edges, it can be considered a *concise GEV network*. Each of these processes results in the removal of nodes or edges from the network, and since any GEV network is finite, the process of reducing any GEV network to its equivalent concise network must conclude after a finite number of transformations. As it is not restrictive to do so, the remainder of this paper will assume that GEV networks are concise.

3.2. Normalization of Logsum Parameters

It is well known that it is necessary to normalize logsum parameters in nested logit models, as the complete set of logsum parameters is over-specified (Ben-Akiva and Lerman, 1985). As the NetGEV model is a generalization of the nested logit model, it follows that the logsum parameters in this model will also need to be normalized. In particular, as mentioned by D&B, the logsum parameters are only relevant in terms of their ratios. This is not quite as obvious in the mathematical formulation presented here as it is in the original formulation, but since they are equivalent the condition still holds. Setting the logsum parameter for any single nest (excepting the nodes associated with elemental alternatives, and degenerate nests) to any positive value will suffice to allow the remaining logsum parameters to be estimated. Typically, it will be convenient to fix the logsum parameter of the root node equal to 1.

The logsum parameters of degenerate nodes (and elemental alternatives) are intrinsically unidentifiable, and thus cannot be used as anchors to identify the parameters on other nodes. If any degenerate node is not removed from the network using Theorem 3.1, then the associated logsum parameter must be set externally.

3.3. Normalization of Allocation Parameters

It is also necessary to normalize the allocation parameters in a NetGEV model. Multiplying all the a values in (2.5) by a constant remains equivalent to multiplying the G function by the constant. More generally, for any network cut that divides the root node from all alternative nodes, multiplying all the a values for all edges in the cut by a constant is equivalent to multiplying G^R by that constant. This change would not affect the ratio of G^R and its derivatives with respect to y, and thus would not affect the resulting probabilities of the model. In order to be able to estimate the allocation parameters, some relationships between them must be fixed externally.

The imposition of these relationships between allocation parameters could potentially create an undesired bias in the model. An unbiased model is one such that the expected value of the random utility for any alternative **i** is equal to the systematic (observed) utility for that alternative, plus a constant with fixed value regardless of the alternative:

$$\bar{U}_{\mathrm{i}} = V_{\mathrm{i}} + \bar{\varepsilon}_{\mathrm{i}} = V_{\mathrm{i}} + \xi, \tag{3.3}$$

and thus $\bar{\varepsilon}_{\mathrm{i}} = \xi$. An unbiased model does not imply that actual observed choice preferences will not be biased in favor of one or more alternatives, but rather indicates merely that a model will not over- or under-predict the probability of an alternative due only to the structure of the model.

The constant expected value of ε, as shown in (3.3), only applies to elemental alternatives. While the log of the generating function G may create a value V which is analagous to the systematic utility of an elemental alternative, there is no explicit error term ε for a nesting node. If one were to be assumed, its expected value could be any value, not necessarily ξ.

The constraint proposed by D&B in equation (52) will not allow the identification of the allocation parameters on edges connected to elemental alternatives. This normalization involves constraining the allocation parameters relative to a ratio of the node-related logsum parameters of the nodes at either end of the associated edge. For elemental alternatives, the μ parameter is intrinsically unidentified, as demonstrated earlier, so imposing a constraint on the allocation parameters based on this parameter will not resolve the identification issue for the allocation parameters.

Instead, Abbé et al. (2007) demonstrate that the proper normalization for a cross-nested logit model (a particular form of a NetGEV with all nesting nodes having the root node as their sole predecessor) is

$$\textstyle\sum_{j \in i^{\uparrow}} \left(a_{ji}\right)^{1/\mu_R} = \kappa \tag{3.4}$$

when the cross-nested network is constructed according to equation (2.5), or

(3.5) $$\sum_{j\in i^{\uparrow}} (a_{ji})^{\mu_j/\mu_R} = \kappa$$

when the cross-nested network is constructed according to equation (2.3), with κ being any constant that does not depend on i or j, typically 1. For convenience, the allocation parameters can be transformed, such that $\alpha_{ij} = a_{ij}^{1/\mu_R}$, resulting in a set of linear constraints on α:

(3.6) $$\sum_{j\in i^{\uparrow}} \alpha_{ji} = \kappa.$$

The Abbe, et al. normalization is not applicable in general to all NetGEV models. Consider the network in Figure 3.1. The marginal distributions of the error terms associated with the three alternatives are:

$$F_{\varepsilon_A}(y_A) = \exp\left[-\exp\left[-\left(y_A - \mu_R \log\left[\alpha_{RK}\alpha_{KA} + \alpha_{RL}\alpha_{LA}\right]\right)\right]\right],$$

$$F_{\varepsilon_B}(y_B) = \exp\left[-\exp\left[-\left(y_B - \mu_R \log\left[\alpha_{RK}\alpha_{KB} + \alpha_{RL}\alpha_{LM}\alpha_{MB}\right]\right)\right]\right],$$

$$F_{\varepsilon_C}(y_C) = \exp\left[-\exp\left[-\left(y_C - \mu_L \log\left[(\alpha_{RL}\alpha_{LC})^{\mu_R/\mu_L} + (\alpha_{RL}\alpha_{LM}\alpha_{MC})^{\mu_R/\mu_L}\right]\right)\right]\right].$$

Under (3.6), the allocation parameters on edges terminating at nodes with a single predecessor would always be κ, in this case that includes α_{RK}, α_{RL} and α_{LM}. This allows a simplification of the marginal distributions:

(3.7) $$F_{\varepsilon_A}(y_A) = \exp\left[-\exp\left[-\left(y_A - \{\mu_R \log\left[\kappa\left(\alpha_{KA} + \alpha_{LA}\right)\right]\}\right)\right]\right],$$

(3.8) $$F_{\varepsilon_B}(y_B) = \exp\left[-\exp\left[-\left(y_B - \{\mu_R \log\left[\kappa\left(\alpha_{KB} + \kappa\alpha_{MB}\right)\right]\}\right)\right]\right],$$

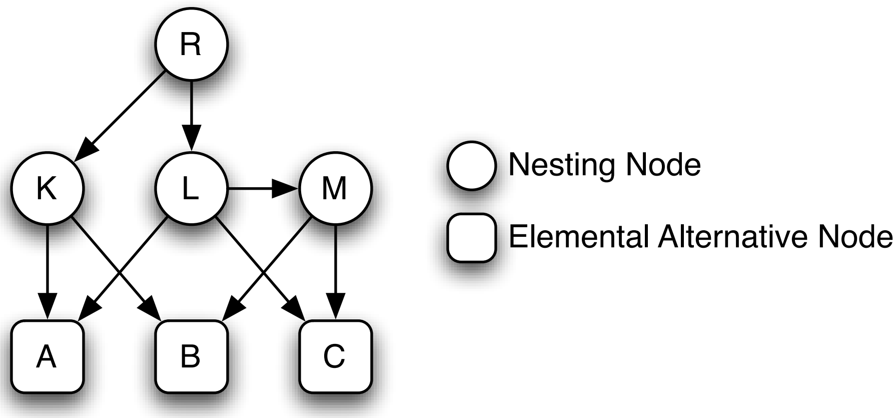


Figure 3.1. A sample network.

$$F_{\varepsilon_C}(y_C) = \exp\left[-\exp\left[-\left(y_C - \left\{\mu_L \log\left[(\kappa\alpha_{LC})^{\mu_R/\mu_L} + \left(\kappa^2\alpha_{MC}\right)^{\mu_R/\mu_L}\right]\right\}\right)\right]\right]. \tag{3.9}$$

Each of (3.7), (3.8) and (3.9) represents an extreme value distribution (as expected for a GEV model) with a scale parameter of 1 and a location parameter indicated by the curled braces. As the scale parameters are all the same, it would be necessary for the location parameters to also be all the same to create an unbiased model. These equations reveal two particular issues in normalizing the allocation parameters for a NetGEV.

First, the possibility of different length paths to elemental alternatives from the root node results in variable size "stack" of allocation values. This can be observed in comparing (3.7) and (3.8), as the longer path R→L→M→B (compared against R→K→A, R→K→B, and R→L→A) adds an extra unbalanced κ in (3.8). This means it is no longer appropriate to normalize using just any constant. Only by setting $\kappa = 1$ will the variance in path lengths become irrelevant, as the edge parameters will not scale against each other. Doing so will cause the location parameters of both (3.7) and (3.8) to collapse to zero, as $\alpha_{KA} + \alpha_{LA} = 1$ and $\alpha_{KB} + \alpha_{MB} = 1$ by (3.6).

The second problem is more complex, as seen in the location parameter in (3.9). What is in essence happening is that the variance of C is decomposed into two parts, which are allocated to the different paths from the root node, in a manner consistent with the decomposition described in Theorem 1 in Abbé et al. (2007). In that decomposition, the error term in the utility of any alternative is represented as the maximum of a group of error terms associated with what might be called "partial alternatives". These partial alternative errors all have a common scale parameter, but various smaller location parameters, such that the maximum of the partial alternative errors is equal to the error of the "whole" alternative.

In the cross-nested model, these partial alternatives recombine at the root node (R) due to the restricted two-level nature of the network, and the errors terms associated with each partial alternative are thus not correlated among each other. In contrast, the two partial alternatives of C in the network in Figure 3.1 "crash" at L, not R. When $\mu_L < \mu_R$, these two parts become positively correlated with each other. The resulting location parameter for the recombined marginal distribution of C will be less than it was before (0), unless $\mu_L = \mu_R$, $\alpha_{LC} = 0$, or $\alpha_{MC} = 0$. One way to ensure that the NetGEV model can be normalized without introducing bias is to constrain the topology of the network to prevent such crashes in the recombination of alternatives.

3.3.1. Crash Free Networks

The restriction necessary to ensure that multiple pieces of the same alternative recombine only at the root node is that all paths to any alternative initially diverge at the root. That is, for any node $i \in \mathcal{C}$, no two distinct paths leading from R to i in may share the edge connected to R. All paths must diverge separately from the root node, and while they may converge sooner than reaching the elemental alternative node, they may not share an edge emanating from the root node and diverge subsequently.

For example, the network on the left side of Figure 3.2 does not conform to this criterion, because elemental alternative C has multiple path divergence points on paths from R. There are four distinct paths through the network from R to C: R $\rightarrow$ M $\rightarrow$ C, R $\rightarrow$ K $\rightarrow$ C, R $\rightarrow$ K $\rightarrow$

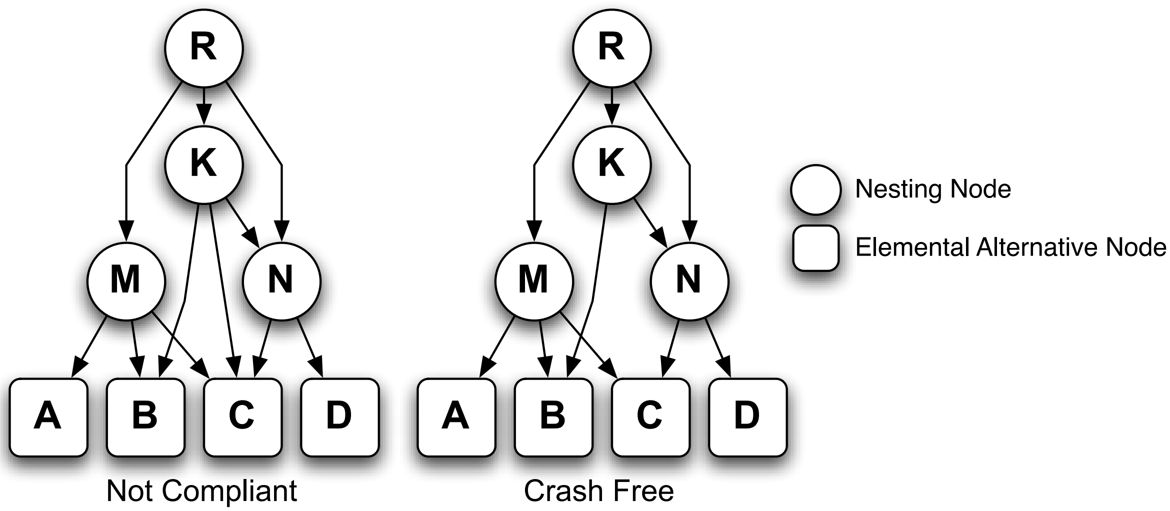


Figure 3.2. Making a GEV network crash-free.

N → C, and R → N → C. The paths R → K → C and R → K → N → C share a common edge emanating from R, which is not allowed. The network on the right side of Figure 3.2 is a different but similar network, with the only difference being that the edge from K to C is missing, eliminating the path R → K → C. Of the three remaining paths, no two share an edge emanating from R. This reduced network is crash free. Note that the crash free network in Figure 3.2 is functionally different from from the original network, and removing an edge from a network can potentially result in a radically different model. (A strategy to adjust a non-conforming network is examined in Section 3.3.4.)

In a crash-free network, for any node except the root node there can be at most one unique path from that node to any other node. If there were more than one path from any node i other than the root node to any other node, then those multiple paths could be extended backwards from i to the root node, sharing common edges, including the edge connecting to the root. Checking this criteria requires building a directed tree from each node connected directly to the root node. If any node in the completed tree has any outbound edges that are not included in the tree, then that edge must connect to another node in the tree, completing a second path to that node, and violating the crash avoidance criterion. Multiple paths diverging from nodes not directly connected to the root node will be captured in the tree[s] of that node's predecessor[s] in the set of nodes connected to the root.

Theorem 3.1. (Crash Free Normalization) *For a GEV network that is crash free, setting the allocation terms* $a_{ij} = \alpha_{ij}^{\mu_R}$ *and enforcing* $\sum_{i \in j^{\uparrow}} \alpha_{ij} = 1$ *will ensure unbiased error terms.*

PROOF. The error terms in the 1-GEV model at any node k are defined by

$$F^k_{\varepsilon_1,...,\varepsilon_J}(y_1, ..., y_J) = \exp\left[-G^k(\exp[-y_1], ..., \exp[-y_J])\right]$$

and, by the requirements for a 1-GEV model, the marginal CDF for any particular error term is

$$F^k_{\varepsilon_{\mathbf{i}}}(y_{\mathbf{i}}) = \exp\left[-\exp\left[-\left(y_{\mathbf{i}} - \log \Theta_{\mathbf{i}}\right)\right]\right],$$

where $\Theta^k_{\mathbf{i}} \equiv G^k(\underset{1}{0}, \ldots, \underset{\mathbf{i}-1}{0}, \underset{\mathbf{i}}{1}, \underset{\mathbf{i}+1}{0}, \ldots, \underset{J}{0})$. This is the extreme value distribution, with location parameter $\log \Theta^k_{\mathbf{i}}$ and a scale parameter of 1. Clearly, the expected value for all the error terms will be equal if and only if $\Theta^R_{\mathbf{i}}$ is equal for all alternatives for the complete model, i.e. at the root node. At the elemental alternative node for node j, $\Theta^j_{\mathbf{j}} = 1$, while $\Theta^j_{\mathbf{i}} = 0$ where $\mathbf{i} \neq \mathbf{j}$. At other nesting nodes $n \in \mathcal{N}$,

$$\Theta^n_{\mathbf{i}} = \left(\sum_{j \in n^{\downarrow}} \left[\left((\alpha_{nj})^{\mu_R}\, \Theta^j_{\mathbf{i}}\right)^{1/\mu_n}\right]\right)^{\mu_n}. \tag{3.10}$$

At the root node,

$$\Theta^R_{\mathbf{i}} = \left(\sum_{n \in R^{\downarrow}} \left[\alpha_{Rn}\left(\Theta^n_{\mathbf{i}}\right)^{1/\mu_R}\right]\right)^{\mu_R}. \tag{3.11}$$

By the definition of a single path divergent network, any node n except the root node must have not more that one unique path to the node i associated with any alternative **i**. Thus, for each such n, at most one term of the summation in (3.10) will have non-zero α, so (3.10) reduces to

$$\Theta_{\mathbf{i}}^{j} = (\alpha_{jk})^{\mu_R}\,\Theta_{\mathbf{i}}^{k}, \tag{3.12}$$

with k being the relevant successor node to j.

For any node n that is an immediate predecessors of i, $\Theta_{\mathbf{i}}^{n} = (\alpha_{ni})^{\mu_R}\,\Theta_{\mathbf{i}}^{i}$. For any node m other than R that is an immediate predecessor of n, $\Theta_{\mathbf{i}}^{m} = (\alpha_{mn})^{\mu_R}\,\Theta_{\mathbf{i}}^{n} = (\alpha_{mn}\alpha_{ni})^{\mu_R}\,\Theta_{\mathbf{i}}^{i}$. This process continues, so that for any node n except R, $\Theta_{\mathbf{i}}^{n}$ is expressed only in terms of α parameters and $\Theta_{\mathbf{i}}^{i}$,

$$\Theta_{\mathbf{i}}^{n} = (\alpha_{nj}\alpha_{jk}\alpha_{kl}\cdots\alpha_{mi})^{\mu_R}\,\Theta_{\mathbf{i}}^{i} = \mathrm{Z}_{ni}^{\mu_R}\Theta_{\mathbf{i}}^{i}, \tag{3.13}$$

where Z_{ni} is the product of all the α parameters associated with all the edges on the unique path through the network from n to i. Substituting (3.13) into (3.11) yields

$$\Theta_{\mathbf{i}}^{R} = \Theta_{\mathbf{i}}^{i}\left(\sum_{n\in R^{\downarrow}}\left[\alpha_{Rn}\mathrm{Z}_{ni}\right]\right)^{\mu_R}. \tag{3.14}$$

At each node j, the allocation parameters subdivide the combined path allocation of all the paths that pass through j to the various predecessor nodes. The requirement that the set of allocation parameters for each node add to 1 ensures that the total incoming allocation equals the total outgoing allocation. There is one path for each term in the summation in (3.14), and the structural restriction ensures that the total allocation among paths equals 1, thus $\sum_{n\in R^{\downarrow}}\left[\alpha_{Rn}\mathrm{Z}_{ni}\right] = 1$, and since $\Theta_{\mathbf{i}}^{i} = 1$ is a constant across all alternatives **i**, $\bar{\varepsilon}_{\mathbf{i}} = \kappa$. □

The crash avoidance restriction is not the only way to allow an unbiased normalization of the allocation parameters in a NetGEV model.

3.3.2. Crash Safe Networks

Crash safe normalization imposes a slightly different restriction on the graph that defines the NetGEV model: for any node $i \in \mathcal{C}$, no two distinct paths leading from R to i may share the edge connected to i. That is, all paths must converge separately at the elemental alternative node, and while they may diverge later than departing the root node, they may not share an edge arriving at the elemental alternative node.

This condition is easier that crash avoidance to check, as only elemental alternative nodes can have multiple predecessor nodes. Since the network is connected and has only one root node without predecessors, every node in the network must have at least one path connecting to it from the root node. If any node j has more than one predecessor node, then it must also have more than one possible path from the root node, as there must be at least one path through each of the predecessor nodes. Those paths would then converge at j. If j is not an elemental alternative node, then the condition for crash safety would be violated.

For example, the network on the left side of Figure 3.3 does not conform to this criterion, because elemental alternative C has multiple path convergence points. There are three distinct paths through the network from R to C: $\mathsf{R} \rightarrow \mathsf{M} \rightarrow \mathsf{C}$, $\mathsf{R} \rightarrow \mathsf{K} \rightarrow \mathsf{M} \rightarrow \mathsf{C}$, and $\mathsf{R} \rightarrow \mathsf{K} \rightarrow \mathsf{N} \rightarrow \mathsf{C}$. The paths $\mathsf{R} \rightarrow \mathsf{M} \rightarrow \mathsf{C}$ and $\mathsf{R} \rightarrow \mathsf{K} \rightarrow \mathsf{M} \rightarrow \mathsf{C}$ share a common edge terminating at C, which is not allowed. The network on the right side of Figure 3.3 is the same, except the edge from K to M is missing, eliminating the path $\mathsf{R} \rightarrow \mathsf{K} \rightarrow \mathsf{M} \rightarrow \mathsf{C}$. The two remaining paths do not share an edge terminating at C. This reduced network is crash safe. Again, the two networks shown in Figure 3.3 represent two different models, with potentially different probabilities for alternatives.

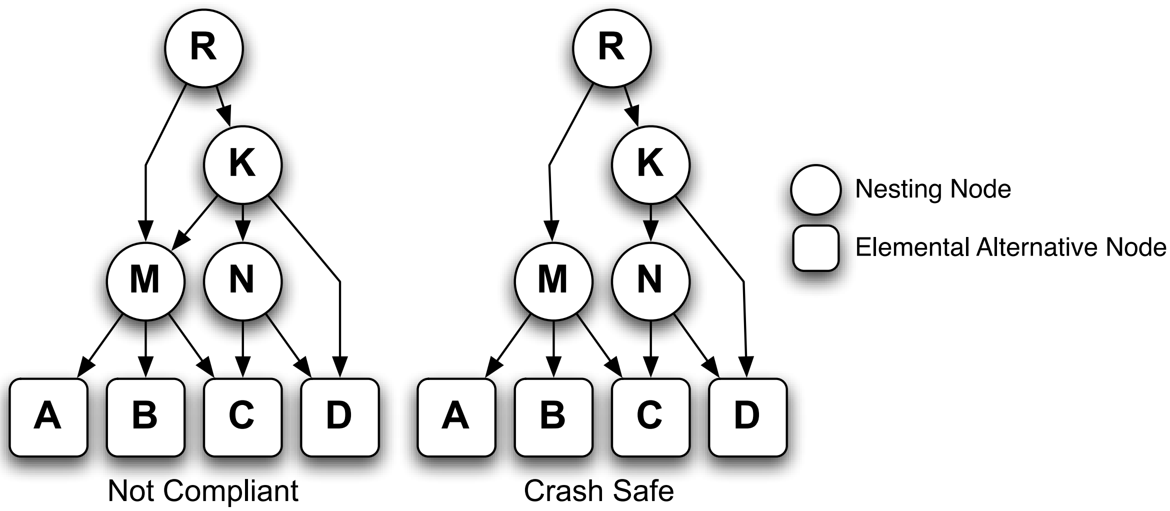


Figure 3.3. Making a GEV network crash safe.

The normalization of a network with this topology is different from that described above. Instead of ensuring that partial allocations of alternatives recombine at the root node (and thus without any internal correlation), the partial alternatives are allowed to recombine at any arbitrary location, with possibly some correlation between the partial alternative's error terms. However, the location of the distribution of the partial alternative error terms is augmented, so that the location of the recombined error distribution will still be constant across alternatives.

In order to provide a general algorithm to ensure this augmentation can be done correctly for each alternative without conflicting with the necessary corrections for other alternatives, all of the splitting of partial alternatives under this topological condition is done on the edges connecting to the elemental alternatives. Each allocation parameter on these edges is associated with one and only one elemental alternative, so that each alternative's partial alternatives can be adjusted independently. It is not necessary that a network is crash safe in this way in order to achieve an unbiased normalized model, if multiple alternatives are constrained such that the necessary adjustments on nesting nodes do not conflict, but it is sufficient and convenient if the criterion described here holds.

The crash safe normalization is more complex that the crash free method, and will require the introduction of some new network descriptors.

As described earlier, each node in $\mathcal{N}$, excluding R, has exactly one predecessor. For any node n in $\mathcal{N}$, let $\dot{n}$ be the predecessor of n, $\ddot{n}$ the predecessor of $\dot{n}$, $\dddot{n}$ the predecessor of $\ddot{n}$, and so on backwards through the network until $\tilde{n}$, which is an eventual predecessor of n and an immediate successor of R.

For each elemental alternative node i, let $\mathcal{G}^i$ be a sub-graph constructed of only the nodes and edges that have i as an eventual successor, excluding i itself.

If $a_{jk} = 1$ for all k in $\mathcal{N}$, then the allocation parameter for the edge connecting from any node in $\mathcal{N}$ to a node i in $\mathcal{C}$ can also be considered as the allocation $\vec{\alpha}_{p_{Ri}}$ to the entire path p_{Ri} from R to i that uses that edge.

For each node j in $\mathcal{N}$, define T(R, j, i) as the set of all paths from R to i that pass through j, and $\tilde{\alpha}_{Rji}$ as the total allocation to those paths:

$$\tilde{\alpha}_{Rji} = \sum_{p \in T(R,j,i)} \vec{\alpha}_{p_{Ri}},$$

or alternatively

$$\tilde{\alpha}_{Rji} = \alpha_{ji} + \sum_{k \in \left\{G^i \cap j^{\downarrow}\right\}} \tilde{\alpha}_{Rki}. \tag{3.15}$$

Theorem 3.2. (Crash Safe Normalization) *For a GEV network which is consistent with single path convergence, setting* $a_{jk} = 1$ *for all* k *in* $\mathcal{N}$ *and*

$$a_{ni} = (\alpha_{ni})^{\mu_n} (\tilde{\alpha}_{Rni})^{\mu_{\dot{n}} - \mu_n} (\tilde{\alpha}_{R\dot{n}i})^{\mu_{\ddot{n}} - \mu_{\dot{n}}} (\tilde{\alpha}_{R\ddot{n}i})^{\mu_{\dddot{n}} - \mu_{\ddot{n}}} \cdots (\tilde{\alpha}_{R\tilde{n}i})^{\mu_R - \mu_{\tilde{n}}}$$

for all i in $\mathcal{C}$, or equivalently,

$$a_{ni} = \left(\frac{\alpha_{ni}}{\tilde{\alpha}_{Rni}}\right)^{\mu_n} \left(\frac{\tilde{\alpha}_{Rni}}{\tilde{\alpha}_{R\dot{n}i}}\right)^{\mu_{\dot{n}}} \left(\frac{\tilde{\alpha}_{R\dot{n}i}}{\tilde{\alpha}_{R\ddot{n}i}}\right)^{\mu_{\ddot{n}}} \left(\frac{\tilde{\alpha}_{R\ddot{n}i}}{\tilde{\alpha}_{R\dddot{n}i}}\right)^{\mu_{\dddot{n}}} \cdots \left(\frac{\tilde{\alpha}_{R\tilde{n}i}}{\tilde{\alpha}_{RRi}}\right)^{\mu_R}, \tag{3.16}$$

and enforcing $\sum_{j \in i^{\uparrow}} \alpha_{ji} = 1$ will ensure unbiased error terms.

Proof. Let $\mathrm{N}_i^{\downarrow}(k)$ be the set of successor nodes to k in $\mathcal{G}^i$. Then for any node $j \in \mathcal{G}^i$,

$$\Theta_{\mathbf{i}}^j = \left(\left(a_{ji}\Theta_{\mathbf{i}}^i\right)^{1/\mu_j} + \sum_{k \in \mathrm{N}_i^{\downarrow}(j)} \left[\left(\Theta_{\mathbf{i}}^k\right)^{1/\mu_j} \right] \right)^{\mu_j}.$$

Let M be a subset of $\mathcal{G}_i$, such that every node m in M has been shown to have $\Theta_{\mathbf{i}}^m$ of the form:

$$\Theta_{\mathbf{i}}^m = \left(\frac{\tilde{\alpha}_{Rmi}}{\tilde{\alpha}_{R\dot{m}i}}\right)^{\mu_{\dot{m}}} \left(\frac{\tilde{\alpha}_{R\dot{m}i}}{\tilde{\alpha}_{R\ddot{m}i}}\right)^{\mu_{\ddot{m}}} \left(\frac{\tilde{\alpha}_{R\ddot{m}i}}{\tilde{\alpha}_{R\dddot{m}i}}\right)^{\mu_{\dddot{m}}} \cdots \left(\frac{\tilde{\alpha}_{R\tilde{m}i}}{\tilde{\alpha}_{RRi}}\right)^{\mu_R} \Theta_{\mathbf{i}}^i. \tag{3.17}$$

Let M' be the compliment of M in $\mathcal{G}^i$. Initially, M is empty. Because $\mathcal{G}^i$ is connected, directed, finite, and circuit-free, there must be at least one node j in M' that has i as its only successor. For any such node, $\Theta_{\mathbf{i}}^j = a_{ji}\Theta_{\mathbf{i}}^i$, so replacing a_{ji} with the right hand side of (3.16) and dropping the first term (as $\alpha_{ni} = \tilde{\alpha}_{Rni}$) yields exactly the form of (3.17). Thus, all such nodes j can be added to M. Consider any node n in M' such that all its successors are in M. Once M has been populated as described above, there must always be at least one such node unless M' is empty. All of the successor nodes m to n are known to have $\Theta_{\mathbf{i}}^m$ of the form in (3.17). With the exception of the first term in the product, the right hand side of (3.17) does not depend on m, so that portion can be brought out of the summation, yielding

$$\Theta_{\mathbf{i}}^n = \left(\frac{\tilde{\alpha}_{Rni}}{\tilde{\alpha}_{R\dot{n}i}}\right)^{\mu_{\dot{n}}} \left(\frac{\tilde{\alpha}_{R\dot{n}i}}{\tilde{\alpha}_{R\ddot{n}i}}\right)^{\mu_{\ddot{n}}} \cdots \left(\frac{\tilde{\alpha}_{R\tilde{n}i}}{\tilde{\alpha}_{RRi}}\right)^{\mu_R} \Theta_{\mathbf{i}}^i \left\{ \frac{\alpha_{ni} + \sum_{m \in \mathrm{N}_i^{\downarrow}(n)} \tilde{\alpha}_{Rmi}}{\tilde{\alpha}_{Rni}} \right\}^{\mu_n}. \tag{3.18}$$

From (3.15), the term in the curly braces in (3.18) equals 1 and can be dropped, leaving the exact form of (3.17), so n can be added to M. Since $\mathcal{G}^i$ is finite, all nodes in $\mathcal{G}^i$ must therefore be in M, including R. $\Theta_{\mathbf{i}}^R$ therefore follows the form of (3.17), so $\Theta_{\mathbf{i}}^R = \Theta_{\mathbf{i}}^i$. As $\Theta_{\mathbf{i}}^i = 1$, which is a constant across all alternatives **i**, so $\bar{\varepsilon}_{\mathbf{i}} = \kappa$. □

3.3.3. Bias Constants

If neither topological condition applies to a GEV network, it is still possible to normalize the allocation parameters and retain an "unbiased" model. One way to do this is to include a complete set of alternative specific constants (except for one arbitrarily fixed reference alternative) in the model. This method does not ensure unbiased systematic utility through constant expected value for the error terms as in (3.3). Instead, $\bar{\varepsilon}_{\mathbf{i}}$ is allowed to vary from κ, but the necessary adjustment $(\bar{\varepsilon}_{\mathbf{i}} - \kappa)$ is incorporated into $V_{\mathbf{i}}$ itself. Unfortunately, this is undesirable because it conflates the model bias correction with the actual choice preference bias. This can cause problems in interpreting these model parameters, and in comparing the parameters between models, even when those models are estimated with the same underlying data. Additionally, there are various reasons why it might be undesirable to include a complete set of alternative specific constants in a model, often because the number of alternatives can be vast for complex models.

3.3.4. Nonlinear Constrained Splitting

If the structure of the GEV network conforms to neither crash free nor crash safe forms, and it is undesirable to include a full set of alternative specific constants, it may still be possible to build an unbiased model through constraints on the form of the allocation values, although these constraints will typically be complex and nonlinear. The easiest way to find the necessary constraints is to decompose the network so that it has the structure needed to apply the crash safe normalizations.

Theorem 3.3. (Node Decomposition) *For any network node $i \in \mathcal{N}$ that has more than one incoming edge (i.e. $\left|i^{\uparrow}\right| = z > 1$), the network can be re-structured by replacing i with z new nodes $i_1, i_2, \ldots, i_z$, each of which has the same μ value and the same set of outgoing edges to successor nodes, but only a single incoming edge from a single predecessor node: $j_1 \rightarrow i_1, j_2 \rightarrow i_2, \ldots j_z \rightarrow i_z$. For each successor node k, the incoming edge from i is replaced with z new incoming edges from $i_1, i_2, \ldots, i_z$. Setting $a_{i_n k_n} = a_{j_n i} a_{i k_n}$ and $a_{j_n i_n} = 1$ for all $n \in \{1, 2, ..., z\}$ will ensure that all nodes in the model excluding i will maintain the same G values, therefore preserving the model probabilities exactly.*

Proof. In the original network, if $i \in j_n^{\downarrow}$ then

$$G_{j_n} = \left(\left\{ a_{j_n i} G^i \right\}^{1/\mu_{j_n}} + \sum_{m \in \{j_n^{\downarrow} \setminus i\}} \left[\left(a_{j_n m} G^m \right)^{1/\mu_{j_n}} \right] \right)^{\mu_{j_n}} . \tag{3.19}$$

In the revised network, the term in the curly braces in (3.19) is replaced by G^{i_n} (as $a_{j_n i_n} = 1$). Multiplying the components of the curly braces yields

$$a_{j_n i} G^i = \left(\sum_{m \in i^{\downarrow}} \left[\left(a_{j_n i} a_{im} G^m \right)^{1/\mu_i} \right] \right)^{\mu_i} = G^{i_n},$$

thus the replacement does not change the value of that term, and hence does not change the value of G^{j_n}. Since this applies for any j_n, the result holds across the entire network. □

Theorem 3.3 can be applied recursively through the network to split any node in $\mathcal{N}$ which has multiple incoming edges. Since $\mathcal{G}$ is circuit free, and the splitting process can only increase the number of incoming edges on successor nodes, the entire network can be restructured to the desired form in a finite number of steps. In each node split, the number of edge allocation values is increased (more edges are added than removed), but the relationship between the allocation values of the additional edges is such that the number of values that can be independently determined remains constant. The final network can then be normalized according to the crash safe

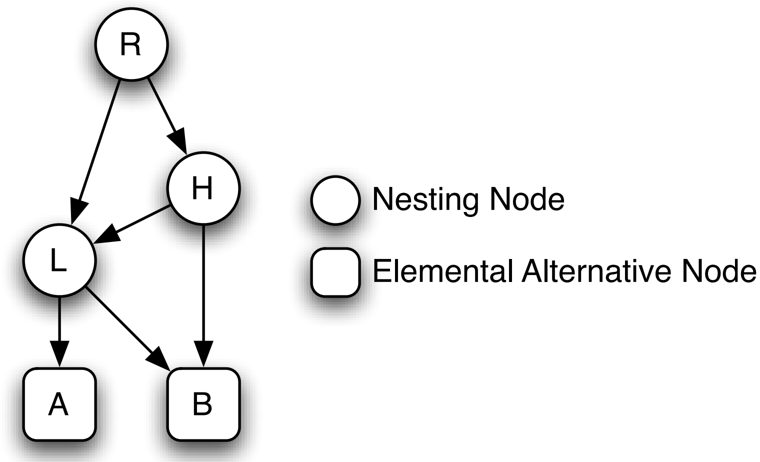


Figure 3.4. A simple network which is neither crash free nor crash safe.

algorithm, subject to the constraints developed in the network decomposition process. A simple network is illustrative of the decomposition process as well as the potential complexity of the non-linear constraints.

For example, consider the simple network depicted in Figure 3.4, which has two elemental alternative nodes, A and B, a root node R, and two other intermediate nesting nodes, H and L. This network conforms to neither the crash free form (R→H→L→B and R→H→B diverge from each other at H, but diverge from R→L→B at R) nor the crash safe form (R→H→L→B and R→L→B converge at L, before converging with R→H→B at B).

The network can be decomposed by splitting L into two new nodes, M and N. One of these nodes inherits the incoming edge from R, while the other inherits the incoming edge from H. Both M and N retain outbound edges to both A and B. The revised network is shown in Figure 3.5.

Unlike the original network in Figure 3.4, the revised network has some constraints imposed on its parameters:

$$\mu_M = \mu_N,$$

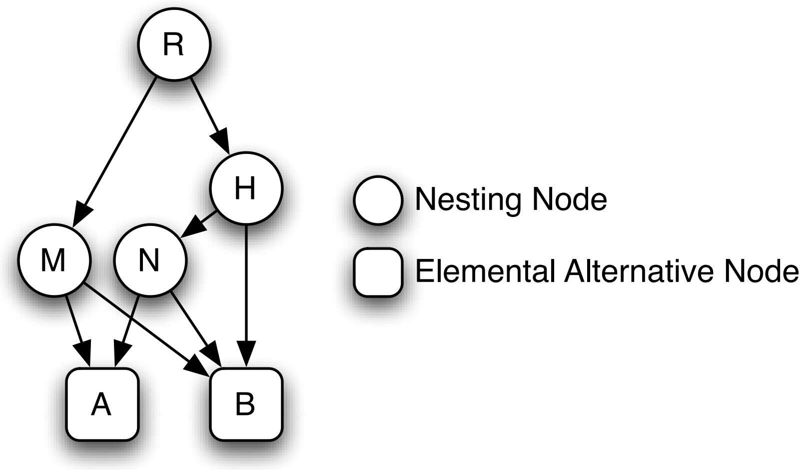


Figure 3.5. A revised network which is crash safe.

$$a_{HN} = 1,$$

$$a_{RM} = 1,$$

(3.20)
$$a_{MA}/a_{NA} = a_{MB}/a_{NB}.$$

The ratio constraint in (3.20) arises from the replacement of a single allocative split at L in Figure 3.4 with two such splits, at M and N, in Figure 3.5. These two splits need to have the same relative ratio, as they are both "controlled" by the ratio of the single split in the original network.

The revised network now meets the structural requirements for crash safe normalization, as only nodes A and B have more than one incoming edge. This normalization replaces the a values with the new values:

$$a_{HB} = \left(\frac{\alpha_{HB}}{\alpha_{HB} + \alpha_{NB}}\right)^{\mu_H} (\alpha_{HB} + \alpha_{NB})^{\mu_R},$$
$$a_{NB} = \left(\frac{\alpha_{NB}}{\alpha_{HB} + \alpha_{NB}}\right)^{\mu_H} (\alpha_{HB} + \alpha_{NB})^{\mu_R},$$
$$a_{MB} = (1 - \alpha_{HB} - \alpha_{NB})^{\mu_R},$$

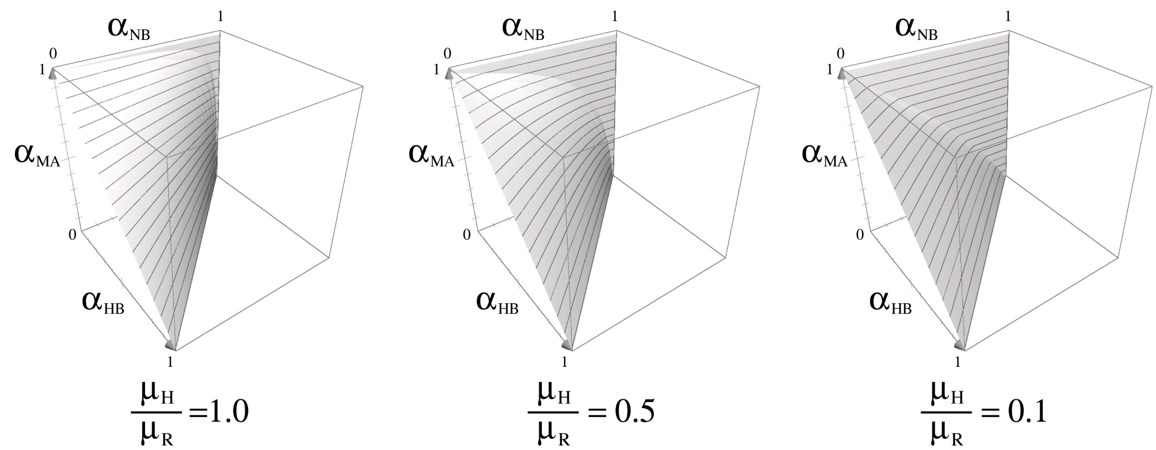


Figure 3.6. Constraint functions for various ratios of μ_H and μ_R.

$$a_{MA} = \alpha_{MA}^{\mu_R},$$

$$a_{NA} = (1 - \alpha_{NA})^{\mu_R}.$$

But from (3.20), we have

$$\alpha_{MA} = \left(\frac{(\alpha_{NB})^{\mu_H/\mu_R} (\alpha_{HB} + \alpha_{NB})^{1-(\mu_H/\mu_R)}}{1 - (\alpha_{HB} + \alpha_{NB})} + 1 \right)^{-1}$$

which is clearly a nonlinear constraint when $0 < \mu_H < \mu_R$.

The shape of the constraint for various different values of μ_H/μ_R is depicted in Figure 3.6. Each constraint surface is depicted inside a unit cube, as each α parameter must fall inside the unit interval, and each surface is defined exclusively in the left triangular region of the cube, because $\alpha_{HB} + \alpha_{NB} \leqslant 1$. In the upper left cube, where $\mu_H/\mu_R = 1.0$, the contour lines of constant α_{MA} are straight, as in that scenario α_{HB} and α_{NB} are linearly related when α_{MA} is otherwise fixed. As μ_H/μ_R approaches 0, the surface of the constraint asymptotically approaches the limiting planes of $\alpha_{MA} + \alpha_{HB} + \alpha_{NB} = 1$ and $\alpha_{HB} = 0$.

CHAPTER 4

Disaggregation of Allocation

4.1. Relaxing Allocation Parameter Constraints

As discussed earlier, the normalization of the NetGEV model requires that the allocation parameters sum to a constant independent of the source node, typically 1. In either the crash safe or crash free conditions, the necessary constraint is $\sum_{j\in i^{\uparrow}} \alpha_{ji} = 1$. Imposing this restriction directly on estimated parameters imposes additional complications, as the parameters are bounded not only by fixed values but also by each other. However, this restriction can be relaxed by transforming the parameters using the familiar logit structure:

$$\alpha_{ji} = \frac{\exp\left(\phi_{ji}\right)}{\sum\limits_{k\in i^{\uparrow}} \left[\exp\left(\phi_{ki}\right)\right]}. \tag{4.1}$$

Under this transformation, a new set of ϕ parameters replaces the α parameters throughout the network on a one-for-one basis. Instead of the α parameters' linear adding-up requirement among the set of parameters associated with each node with more than one predecessor, the ϕ parameters may vary unbounded across $\mathbb{R}$, so long as one ϕ in each such group is fixed to some constant value (typically zero). This is a significant advantage in parameter estimation, as nonlinear optimization algorithms are substantially easier to implement when there are no (or fewer) constraints on the parameters.

4.2. Subparameterization of Allocation

Replacing the α parameters with a logit formulation not only simplifies the process of estimating the allocation parameters, it also opens up the possibility creating a much richer model.

The logit structure for nest allocation allows for the incorporation of data into the correlation structure of error terms:

$$\alpha_{tji} = \frac{\exp\left(\phi^{\star}_{ji} + \phi_{ji} Z_t\right)}{\sum\limits_{k \in i^{\uparrow}} \left[\exp\left(\phi^{\star}_{ki} + \phi_{ki} Z_t\right)\right]}, \tag{4.2}$$

where $\phi^{\star}_{ji}$ is the baseline parameter as in (4.1), Z_t is a vector of data specific to decision maker t, and ϕ_{ji} is a vector of parameters to the model which are specific to the link from predecessor node j to successor node i. If we assume that the first value in Z_t is 1, we can simplify (4.2) to

$$\alpha_{tji} = \frac{\exp\left(\phi_{ji} Z_t\right)}{\sum\limits_{k \in i^{\uparrow}} \left[\exp\left(\phi_{ki} Z_t\right)\right]}. \tag{4.3}$$

Thus the G function for nesting nodes becomes

$$G^i(y) = \left(\sum_{j \in i^{\downarrow}} \left(\frac{\exp\left(\phi_{ji} Z_t\right)}{\sum\limits_{k \in i^{\uparrow}} \left[\exp\left(\phi_{ki} Z_t\right)\right]} G^j(y) \right)^{1/\mu_i} \right)^{\mu_i}.$$

The ϕ parameters are all arc specific parameters, analogous to alternative specific parameters in an MNL model. As usual for "alternative" specific constants and variables logit models, one of the vectors ϕ_{ji} must be constrained to some arbitrary value, usually zeros. The remaining ϕ vectors can vary unconstrained in both positive and negative regions of $\mathbb{R}$. This formulation also allows the addition of decision-maker attributes to be introduced as data to the model, not only in determining the systematic (observed) utility, but also in determining the correlation structure for random (unobserved) utility.

This new mathematical form, a heterogeneous covariance network generalized extreme value model (HeNGEV), is substantially different from any closed form model found in the literature. Unlike the heterogeneous correlation nested model proposed by Bhat (1997a), which allows heterogeneous covariance values but requires a single covariance structure among alternatives, the

HeNGEV model allows both the amount and the form of covariance to vary across decision makers.

For example, consider a network model of itinerary and fare class choice bifurcated into two sub-structures, one with itinerary nested inside fare class, and the other with fare class nested inside itinerary. The allocation parameters could then vary based on frequent flier status, with program member decision makers tending to choose based on one substructure, and nonmember decision makers tending to choose based on the other.

Since the form of (4.3) is by construction strictly positive, the HeNGEV model inherits consistency with utility maximization from the NetGEV formulation, as long as the same conditions hold (in particular, non-increasing node parameters in the network). Achieving covariance heterogeneity through the allocation parameters is simpler than subparameterization of the node parameters, as in Bhat (1997a). When nodes have a single parameter, as in a nested logit model, the decreasing node parameter condition can be enforced with a single simple inequality condition on each node (comparing it to its single predecessor). In a NetGEV model, on the other hand, the decreasing node parameter condition requires a minimization process in addition to an inequality (a node parameter must be smaller than the smallest of its predecessors). Thus, achieving such covariance heterogeneity in a NetGEV model using the node parameters would result in a model with a non-differentiable likelihood function, and maximum likelihood estimation would be much more difficult.

CHAPTER 5

Derivatives and Elasticity

Because of the flexible nature of the HeNGEV model, it is possible to create structures with complex or unusual cross-elasticities. Unlike the MNL and NL models, representing the derivatives and elasticities for HeNGEV models in general is inconvenient to do in a simple formula. Instead, the they can be understood by mathematically localizing the derivatives, elasticities, probabilities, and utilities withing the generating network.

5.1. Derivatives of Log Likelihood with Respect to Parameters

The derivatives of the log likelihood with respect to the parameters, just as in other GEV models, is

$$\frac{\partial LL}{\partial \Xi} = \sum_{t} \sum_{i \in C} \left[\delta_{ti} \frac{1}{P_{ti}} \frac{\partial P_{ti}}{\partial \Xi} \right],$$

with t indexing across decision makers, and Ξ as a vector of all model parameters. The derivative of the log likelihood is thus a function of the derivative of the probabilities,

$$\frac{\partial P_{ti}}{\partial \Xi} = \sum_{k \in i^{\uparrow}} \left[P_{ti|k} \left(\frac{\partial P_{tk}}{\partial \Xi} + P_{tk} Q_{\Xi tki} \right) \right] \tag{5.1}$$

with

$$Q_{\Xi tki} = \frac{1}{\mu_k} \left(\frac{\partial V_{ti}}{\partial \Xi} - \frac{\partial V_{tk}}{\partial \Xi} + \frac{V_{tk} - V_{ti} - \log \alpha_{tki}}{\mu_k} \frac{\partial \mu_k}{\partial \Xi} + Z_t \cdot \frac{\partial \phi_{ki}}{\partial \Xi} - \sum_{m \in i^{\uparrow}} \left[\alpha_{tmi} \left(Z_{tb} \cdot \frac{\partial \phi_{mib}}{\partial \Xi} \right) \right] \right)$$

representing the network-localized perturbation of the probability for t along the edge from k to i. Q can be simplified for each of the various parameter types:

$$Q_{\beta tki} = \frac{1}{\mu_k}\left(\frac{\partial V_{ti}}{\partial \Xi} - \frac{\partial V_{tk}}{\partial \Xi}\right), \tag{5.2}$$

$$Q_{\mu tki} = \frac{1}{\mu_k}\left(\frac{\partial V_{ti}}{\partial \Xi} - \frac{\partial V_{tk}}{\partial \Xi} + \left(\frac{V_{tk} - V_{ti} - \log \alpha_{tki}}{\mu_k}\right)\frac{\partial \mu_k}{\partial \Xi}\right), \tag{5.3}$$

and

$$Q_{\phi tki} = \frac{1}{\mu_k}\left(\frac{\partial V_{ti}}{\partial \Xi} - \frac{\partial V_{tk}}{\partial \Xi} + \left(Z_t \cdot \frac{\partial \phi_{ki}}{\partial \Xi}\right) - \sum_{m \in i^{\uparrow}}\left[\alpha_{tmi}\left(Z_{tb} \cdot \frac{\partial \phi_{mib}}{\partial \Xi}\right)\right]\right). \tag{5.4}$$

Having isolated the localized perturbation in Q, the remaining portion of (5.1) serves to stack that local perturbation with the derivatives of probability at the next higher level of nesting. Expanding (5.1) at the next higher level yields

$$\begin{aligned}
\frac{\partial P_{ti}}{\partial \Xi} &= \sum_{k \in i^{\uparrow}}\left[P_{ti|k}\left(\left\{\sum_{j \in k^{\uparrow}}\left[P_{tk|j}\left(\frac{\partial P_{tj}}{\partial \Xi} + P_{tj}Q_{\Xi tjk}\right)\right]\right\} + P_{tk}Q_{\Xi tki}\right)\right] \\
&= \sum_{k \in i^{\uparrow}}\left[\left(\sum_{j \in k^{\uparrow}}\left[\left(P_{ti|k}P_{tk|j}\frac{\partial P_{tj}}{\partial \Xi} + P_{ti|k}P_{tk|j}P_{tj}Q_{\Xi tjk}\right)\right] + P_{ti|k}P_{tk}Q_{\Xi tki}\right)\right] \\
&= \sum_{k \in i^{\uparrow}}\left[\left(\sum_{j \in k^{\uparrow}}\left[\left(P_{ti|k}P_{tk|j}\frac{\partial P_{tj}}{\partial \Xi} + P_{ti|k}P_{tk|j}P_{tj}Q_{\Xi tjk}\right)\right] + P_{ti|k}\sum_{j \in k^{\uparrow}}\left[P_{tk|j}P_{tj}\right]Q_{\Xi tki}\right)\right] \\
&= \sum_{k \in i^{\uparrow}}\sum_{j \in k^{\uparrow}}\left[P_{ti|k}P_{tk|j}\frac{\partial P_{tj}}{\partial \Xi} + P_{ti|k}P_{tk|j}P_{tj}\left(Q_{\Xi tjk} + Q_{\Xi tki}\right)\right]. \qquad (5.5)
\end{aligned}$$

This expansion can be repeated all the way through the network, following the form of (5.5), until reaching the root node, which has constant probability of 1 (and a derivative of 0), thus eliminating the front term in (5.5). So the derivative of probability of an alternative with respect to the parameters becomes

$$\frac{\partial P_{ti}}{\partial \Xi} = \sum_{\vec{p}_{Ri} \in T(R,i)} \left[P_{ti\vec{p}_{Ri}} \sum_{jk \in \vec{p}_{Ri}} Q_{\Xi tjk} \right],$$

the sum of arc effects on each path, multiplied by the path probability, summed over all paths.

5.2. Derivatives of Utility with Respect to Parameters

The arc effects on probability Q shown in (5.2-5.4) vary depending on the parameter type, but all rely in part on the derivatives of the systematic utility with respect to the parameters. For each of the different parameter types in the model, the derivatives of utility can be subdivided into two classes: direct and indirect effects. Direct effects of a change in a parameter occur in those nodes where the parameter appears specifically in the generation of that node's utility. The set of nodes that incur direct effects in utility for a parameter Ξ_b are designated D_{Ξ_b}. Other nodes (the compliment of D_{Ξ_b} in $\mathcal{Z}$) are in the set $\bar{D}_{\Xi_b}$.

For β parameters, direct effects occur only in the elemental alternative nodes, so $D_\beta = \mathcal{C}$. The derivative of utility at those nodes, when utility is defined as a linear in parameters function, is merely the data:

$$\frac{\partial V_{ti}}{\partial \beta_b} = X_{tib},\ i \in D_{\beta_b}.$$

For μ parameters, the direct effects occur only in the node associated with each individual μ, so that $D_{\mu_i} = \{i\}$, and

$$\frac{\partial V_{ti}}{\partial \mu_i} = \frac{1}{\mu_i} \left(V_{ti} - \sum_{j \in i^\downarrow} \left[P_{tj|i} \left(V_{tj} + \log \left[\alpha_{tij} \right] \right) \right] \right).$$

For ϕ, the direct effects occur only in the set of predecessor nodes associated the the set of ϕ's, such that $D_{\phi_{ik}} = k^\uparrow$. However, the form of the derivative of V_{ti} with respect to ϕ_{jk} depends on whether $i = j$, i.e. whether i is the source of the edge associated with the ϕ. When $i = j$, then

$$\frac{\partial V_{ti}}{\partial \phi_{ik}} = P_{tk|i}(1 - \alpha_{tik})Z_t.$$

When $i \neq j$ but $i \in k^{\uparrow}$,

$$\frac{\partial V_{ti}}{\partial \phi_{jk}} = P_{tk|i}\left(-\alpha_{tjk}\right)Z_t.$$

Indirect effects, as opposed to direct effects, occur when the utility of a node changes because the utility of one of its successor nodes changes. Due to the formulation of the model, indirect effects can only propagate upwards through the network (towards the root node, away from the elemental alternatives). Once the impact on utility has been isolated from the initial cause, the ripple effects upward through the network no longer rely on the parameter type or the location of the source:

$$\frac{\partial V_{ti}}{\partial \Xi_b} = \sum_{j \in i^{\downarrow}} \left[P_{tj|i} \left(\frac{\partial V_{tj}}{\partial \Xi} \right) \right], \forall i \in \bar{D}_{\Xi_b}.$$

5.3. Elasticity with Respect to Utility Variables

Similar the the derivatives with respect to the parameters, elasticity with respect to variables is best understood as a sequential layering of effects. Elasticity is a function of the probability, and probability is a function of utility. The utility of the elemental alternatives forms the basic foundation of the model, generally expressed as the linear form $V_{ti} = X_{ti}\beta$. From that foundation, the utility of other nodes in the generating network is determined, with the utility of each node determined only by the utilities of its immediate downstream neighbors. The determination of utility thus moves upstream from the elemental alternatives up through the network, to the root node. Then, conditional and unconditional probabilities for nodes can be calculated, with each node allocating its probability to its immediate downstream neighbors, based on the utilities of those neighbors. The probability cascades downstream from the root node, which by definition always has a probability of 1.

The formula for point elasticity of demand with respect to attributes of the alternatives (X_{tj}) is

$$\eta_{ti,x_{tj}} = \frac{\partial P_{ti}}{\partial X_{tj}} \frac{X_{tj}}{P_{ti}} \tag{5.6}$$

In most logit models, this formula is usually partitioned, using the chain rule, to be

$$\eta_{ti,x_{tj}} = \frac{\partial P_{ti}}{\partial V_{tj}} \frac{\partial V_{tj}}{\partial X_{tj}} \frac{X_{tj}}{P_{ti}}.$$

This partition is useful when V_{tj} is a linear in parameters function (i.e. $V_{tj} = X_{tj}\beta$), as that means $\partial V_{tj}/\partial X_{tjk} = \beta_k$, so that the complexity of the formulation is quarantined inside $\partial P_{ti}/\partial V_{tj}$.

However, if P_{tj} in (5.6) is expanded instead of applying the chain rule, the derivative in the elasticity formulation develops a structure that mirrors the natural partitioned mathematical structure of the network:

$$\frac{\partial P_{ti}}{\partial X_{tjk}} = \frac{\partial}{\partial X_{tjk}} \sum_{n \in i^{\uparrow}} \left[P_{tk} \left(\frac{\alpha_{tni} G_{ti}}{G_{tn}} \right)^{1/\mu_n} \right] = \sum_{n \in i^{\uparrow}} \left[P_{ti|n} \left(\frac{\partial P_{tn}}{\partial X_{tjk}} + \frac{P_{tn}}{\mu_n} \left(\frac{\partial V_{ti}}{\partial X_{tjk}} - \frac{\partial V_{tn}}{\partial X_{tjk}} \right) \right) \right]$$

so that the partial derivatives of probability at any particular node are a function of the partial derivatives of probability at adjacent upstream nodes, as well as the partial derivatives of utility at the node and its upstream neighbors. The component partial derivatives of probability flow downstream from earlier nodes, just as probability flows downstream, with those derivatives anchored at the root node, where the derivative of probability is always 0 (since the probability at the root is always 1).

This form then relies not merely on the derivative of utility at the alternative, but also on the derivatives of utility through the entire network. Like the derivatives of probability, which match the original flow of probability down from the root node, the derivatives of utility match the flow of utility up from the elemental alternatives. The component partial derivatives of V are

$$\frac{\partial V_{ti}}{\partial X_{tjk}} = \begin{cases} \beta_k & i = j,\ i \in \mathcal{C} \\ 0 & i \neq j,\ i \in \mathcal{C},\ j \in \mathcal{C} \\ \sum_{n \in i\downarrow} \left[P_{tn|i} \left(\partial V_{tn} / \partial X_{tj} \right) \right] & i \in \mathcal{N},\ j \in \mathcal{C} \end{cases}$$

Critically, these derivatives depend on probability but not the derivatives of probability. This allows a sequential process of calculations: first utility flows up, then probability flows down, then the derivatives of utility flow up, and finally the derivatives of probability can flow down through the network.

5.4. Elasticity with Respect to Allocative Variables

When a HeNGEV model is created with data entering the allocation terms as well as the utility terms, it is also useful to examine the elasticity with respect to this second data pool. The elasticity with respect to allocation follows a fairly similar functional form, as the propagating indirect effects of a change in the data are transmitted through the network in the same manner, through the utility and probability. It is only in the direct effects on individual nodes that the elasticity of allocation variables differs from the elasticity of utility variables.

The effect of allocative variables on the systematic utility of elemental alternative nodes is always zero. Since the allocation parameters only appear in the formula for nesting nodes, they can have a direct affect only on those nodes. Above the elemental alternatives, the utility of nodes is responsive not only to indirect changes in utility through changes in immediate successors, but also to direct changes induced in the allocations on the attached edges,

$$\frac{\partial V_{ti}}{\partial Z_{tb}} = \begin{cases} 0 & i \in \mathcal{C} \\ \sum\limits_{j \in i^{\downarrow}} \left[P_{tj|i} \left(\frac{\partial V_{tj}}{\partial Z_{tb}} + \phi_{ijb} - \sum\limits_{n \in j^{\uparrow}} \left[\alpha_{tnj} \phi_{njb} \right] \right) \right] & i \in \mathcal{N} \end{cases} .$$

The indirect effects on utility propagate toward the root node, just as for utility variables. The derivatives of probability similarly are responsive to changes in utility, as well as changes in attached allocations,

$$\frac{\partial P_{ti}}{\partial Z_{tb}} = \sum_{n \in i^{\uparrow}} \left[P_{ti|n} \left(\frac{\partial P_{tn}}{\partial Z_{tb}} + \frac{P_{tn}}{\mu_n} \left(\phi_{nib} - \sum_{m \in i^{\uparrow}} \left[\alpha_{tmi} \phi_{mib} \right] + \frac{\partial V_{ti}}{\partial Z_{tb}} - \frac{\partial V_{tn}}{\partial Z_{tb}} \right) \right) \right] .$$

Just as for utility variables, the derivatives of probability at particular nodes are determined only by the derivatives of utility and the derivatives of probability for immediate predecessor nodes.

CHAPTER 6

Application

The HeNGEV model, by its nature, is resistant to testing using very simple data sets. Various simple HeNGEV mode choice models were explored for work and shopping trips in the San Francisco metropolitan area, using small publicly available data sets published by the Metropolitan Transportation Commission, but none provided a valid model with parameters consistent with the necessary conditions for a NetGEV model. Instead, the benefits of complex heterogeneous correlation structures are most likely to be visible when choice sets are larger, and correlations are likely to occur across many dimensions. Rather than exploring many more data sets to find one that would reveal such a correlation structure, the model can be tested using synthetically generated data. This provides a good test to see if the HeNGEV model form can correctly recover the model parameters that the homogeneous NetGEV form cannot, when the observed choices are known *a priori* to have been made in a manner consistent with the HeNGEV structure and a particular known set of parameters.

6.1. Data Generation

A data set was generated that would approximate the data that might be observed for a flight itinerary choice. The flight itinerary choices were based on observed flight itineraries in 2001 for travel from Cincinnati-Northern Kentucky International Airport (CVG) to Albuquerque International Sunport (ABQ), using data extracted from Coldren (2004). This airport pair was used because it provided a variety of itinerary options (non-stop, single connect, and double connect flights on five different carriers) within a relatively small number of total possible itineraries (28 distinct itineraries). This allowed the creation of an interesting structural network of choices, without the total number of available choices being overwhelmingly large. A much larger set of alternatives should not cause behavior to be fundamentally different, but would increase computing time to estimate the models in this experiment. From these observed itineraries, various data attributes were used, including departure time, level of service (non-stop, single connect, double connect), carrier (American, Continental, Delta, Northwest, United), fare ratio (the comparative fare levels, on average, across the airlines serving this city pair), and distance ratio (the ratio of itinerary flight distance to straight line distance). The data on the itineraries is shown in Table 6.1.

The advantage of the HeNGEV model introduced in this dissertation is that it can incorporate attributes of the decision maker (or of the choice itself) into the correlation structure. To examine the usefulness of such enhanced tools, the synthetic data set also included data on the annual income level of each decision maker, as well as the number of days in advance that the ticket was purchased. Income was randomly generated for each traveler using a polynomial distribution, with values ranging from \$30 thousand to \$180 thousand. Advance purchase times were generated using a uniform distribution ranging between 0 and 28 days, with a partial negative correlation to income (i.e. higher income tended to match with shorter advance purchase time frames). These attributes were generated for 100,000 simulated travelers.

Once the explanatory variables were created, a plausible HeNGEV choice model was created. The structure of this model is depicted in Figure 6.1. The network depicted has numerous nodes and arcs. If the associated parameters were each estimated independently, the parameter estimation process would become overwhelmed, and the resulting model would be virtually meaningless as a descriptive or predictive tool. Instead, the nodes are grouped into four sections (upper and

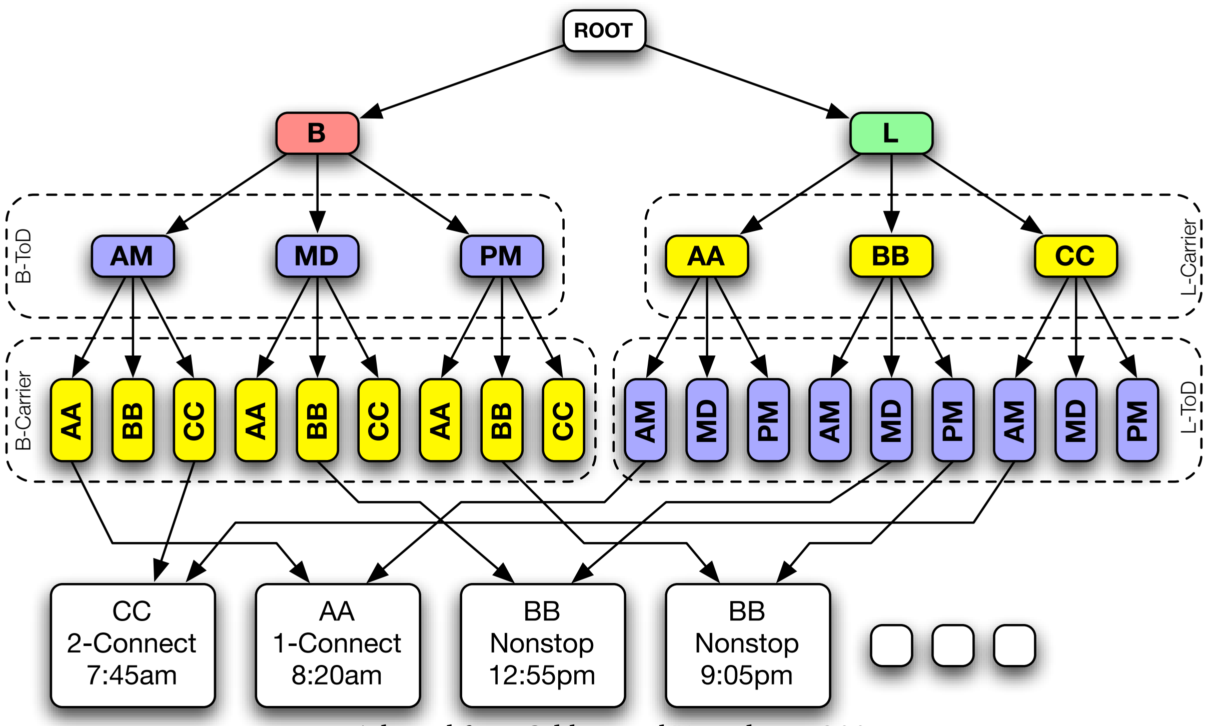
Table 6.1. Flight Itinerary Choices in Synthetic Data

Itinerary Number	**Airline**	**Departure Time**	**Distance Ratio**	**Fare Ratio**	**Level of Service**
1	BB	12:55	100	104	Non-stop
2	BB	21:05	100	104	Non-stop
3	AA	13:19	111	100	Single Connect
4	AA	16:47	111	100	Single Connect
5	AA	16:47	111	100	Single Connect
6	AA	8:20	111	100	Single Connect
7	AA	16:15	111	100	Single Connect
8	CC	18:20	127	55	Single Connect
9	CC	9:15	127	55	Single Connect
10	BB	16:45	132	104	Single Connect
11	BB	14:50	132	104	Single Connect
12	BB	7:20	132	104	Single Connect
13	BB	12:30	111	104	Single Connect
14	BB	17:05	111	104	Single Connect
15	BB	18:50	111	104	Single Connect
16	BB	7:45	111	104	Single Connect
17	DD	9:15	127	46	Single Connect
18	DD	18:20	127	46	Single Connect
19	CC	8:00	130	55	Single Connect
20	BB	9:00	132	104	Single Connect
21	AA	10:05	132	100	Double Connect
22	AA	16:15	132	100	Double Connect
23	AA	14:40	132	100	Double Connect
24	BB	11:00	153	104	Double Connect
25	DD	7:15	130	46	Double Connect
26	DD	14:40	130	46	Double Connect
27	EE	7:30	121	49	Double Connect
28	EE	7:30	121	49	Double Connect

lower nests on each side) with common logsum parameters, and the allocations between the sides were grouped together so that all alternatives would have common allocation parameters.

A choice model was created, with utility parameters drawn from similar models estimated by Coldren (2004) (see also Coldren and Koppelman, 2005). No similar structure model had been estimated for the correlation across time of day and carrier, so values for those logsum parameters were approximated. No model has ever been estimated with allocation parameters on data in the

Figure 6.1. Flight Itinerary Choice Model for Synthetic Data



Adapted from Coldren and Koppelman, 2005

HeNGEV model, so values for those parameters were selected such that the range of observed data for income and advanced purchase would result in a broad range of allocation values across the sample, with large densities near the extreme values (i.e. numerous simulated travelers choose based almost exclusively on one sub-model or the other). The distribution of allocations for the simulated travelers, based on the generating model, is shown in Figure 6.2.

Such an extreme bimodal distribution of the error covariance structure might not be often observed in actual data, so the performance of the HeNGEV model was also tested using a second dataset with a more unequal distribution of error covariance structures. This second dataset was generated using all the same input values and utility parameters, but changing only the coefficient on income in the allocation function, from -0.03 to -0.04. This small change created a large skew in the resulting allocations for simulated travelers, as seen in Figure 6.3.

Figure 6.2. Distribution of Allocation Weights in Bimodal Synthetic Data

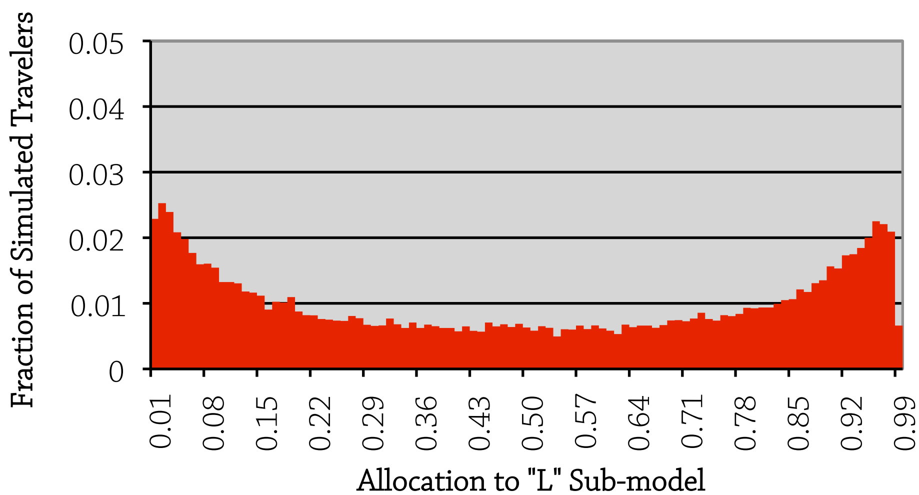
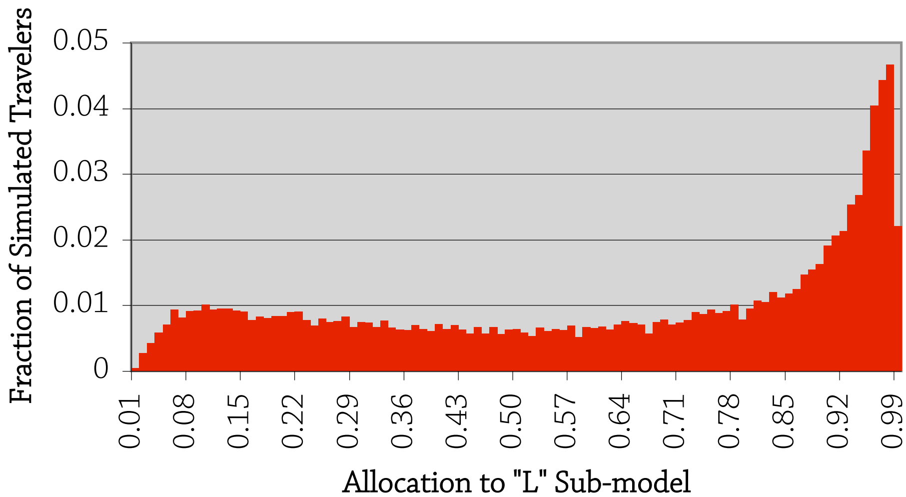


Figure 6.3. Distribution of Allocation Weights in Unimodal Synthetic Data



6.2. Results

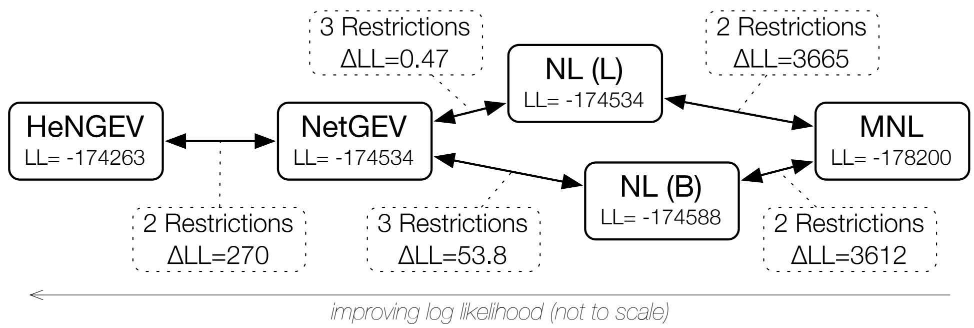
Overall, the HeNGEV model performs better than the matching NetGEV formulation in predicting the observed simulated choices, in both the bimodal and unimodal synthetic datasets.

6.2.1. Bimodal Dataset

The estimated parameters for the HeNGEV model for the bimodal dataset are shown in Table 6.2.1. None of the estimated parameters in this model differs from the known true parameters by a statistically significant amount (although the lower level logsum parameter in the L sub-model comes close to a statistically significant deviation).

The HeNGEV model can be compared to the NetGEV model which matches the same network structure, but lacks the heterogeneous covariance made possible by the disaggregation of the allocation parameters. The estimated parameters for the matching NetGEV model are shown in Table 6.3.The NetGEV model also performs somewhat well, estimating most of the parameters near their known true values. However, the performance of the HeNGEV model is generally

Figure 6.1. Log Likelihoods and Relationships Between Models Estimated Using Bimodal Dataset



superior to the NetGEV model, as shown in Table 6.4. Of 13 utility and logsum parameters in the models, 11 are closer to their true values in the HeNGEV model than in the NetGEV model, mostly by more than half the NetGEV's error. The overall log likelihood is also much better for the HeNGEV model (270 higher than the NetGEV model), a statistically highly significant difference for two degrees of freedom.

The utility function used to create these models can also be used in regular nested logit models (including only one of the submodels) or a regular multinomial logit model. The various results for these models are summarized in Table 6.5. A graphical representation of the relationship between the various estimated models for the bimodal data is shown in Figure 6.1. In this dataset, the nested logit models fit the data nearly as well as the NetGEV model. The difference between the NetGEV and L-only model, only 0.47 in log likelihood, is not statistically significant (χ^2=0.94, with 3 degrees of freedom, p=0.82). The B-only model fits somewhat worse, but returns parameter estimates for the logsum parameters that are out of the acceptable range.

Table 6.2. HeNGEV Model for Bimodal Dataset

Log Likelihood at Convergence:		-174263.42		
Log Likelihood at Zeros:		-333220.45		
Rho Squared w.r.t. Zeros:		0.477		
	True Value	**Estimated Parameter**	**Std. Error of Estimate**	**t Statistic vs. True**
Departure Time				
Before 08:00	0	0	n/a	n/a
08:00-09:59	0.15	0.1775	0.0164	1.68
10:00-12:59	0.1	0.01935	0.1266	-0.64
13:00-15:59	0.05	0.08199	0.02451	1.31
16:00-18:59	0.1	0.1058	0.01836	0.32
19:00 or later	-0.3	-0.3872	0.1264	-0.69
Level of Service				
Nonstop	0	0	n/a	n/a
Single Connect	-2.3	-2.415	0.1287	-0.89
Double Connect	-5.8	-5.959	0.1588	-1.00
Flight Characteristics				
Distance Ratio	-0.01	-0.01159	0.0008905	-1.79
Fare Ratio	-0.004	-0.004766	0.0004494	-1.70
Nesting Parameters				
B Time of Day (Upper) Nest	0.8	0.8034	0.01364	0.25
B Carrier (Lower) Nest	0.2	0.2277	0.01736	1.60
L Carrier (Upper) Nest	0.7	0.7139	0.01714	0.81
L Time of Day (Lower) Nest	0.3	0.3144	0.007438	1.94
Allocation Parameters				
Phi Constant L Side	1	1.05	0.499	0.10
Phi Income (000) L Side	-0.04	-0.04159	0.007447	-0.21
Phi Advance Purchase L Side	0.2	0.2185	0.03331	0.56

Table 6.3. NetGEV Model for Bimodal Dataset

Log Likelihood at Convergence:		-174533.95		
Log Likelihood at Zeros:		-333220.45		
Rho Squared w.r.t. Zeros:		0.476		

	True Value	**Estimated Parameter**	**Std. Error of Estimate**	**t Statistic vs. True**
Departure Time				
Before 08:00	0	0	n/a	n/a
08:00-09:59	0.15	0.2122	0.0208	2.99
10:00-12:59	0.1	0.05807	0.1495	-0.28
13:00-15:59	0.05	0.09039	0.09349	0.43
16:00-18:59	0.1	0.1213	0.06865	0.31
19:00 or later	-0.3	-0.3594	0.1631	-0.36
Level of Service				
Nonstop	0	0	n/a	n/a
Single Connect	-2.3	-2.431	0.2356	-0.56
Double Connect	-5.8	-6.013	0.4957	-0.43
Flight Characteristics				
Distance Ratio	-0.01	-0.01373	0.001208	-3.09
Fare Ratio	-0.004	-0.005724	0.0006717	-2.57
Nesting Parameters				
B Time of Day (Upper) Nest	0.8	0.8349	0.2855	0.12
B Carrier (Lower) Nest	0.2	0.3061	0.03978	2.67
L Carrier (Upper) Nest	0.7	0.7471	0.4955	0.10
L Time of Day (Lower) Nest	0.3	0.2776	0.118	-0.19
Allocation Parameters				
Phi Constant L Side	1	0.1039	3.051	-0.29

Table 6.4. Comparison of NetGEV and HeNGEV Models for Bimodal Data

	HeNGEV Model		NetGEV Model	
	Actual Error of Estimate	**Standard Error of Estimate**	**Actual Error of Estimate**	**Standard Error of Estimate**
Departure Time				
Before 08:00	n/a	n/a	n/a	n/a
08:00-09:59	0.0275	0.0164	0.0622	0.0208
10:00-12:59	-0.08065	0.1266	-0.04193	0.1495
13:00-15:59	0.03199	0.02451	0.04039	0.09349
16:00-18:59	0.0058	0.01836	0.0213	0.06865
19:00 or later	-0.0872	0.1264	-0.0594	0.1631
Level of Service				
Nonstop	n/a	n/a	n/a	n/a
Single Connect	-0.115	0.1287	-0.131	0.2356
Double Connect	-0.159	0.1588	-0.213	0.4957
Flight Characteristics				
Distance Ratio	-0.00159	0.0008905	-0.00373	0.001208
Fare Ratio	-0.000766	0.0004494	-0.001724	0.0006717
Nesting Parameters				
B Time of Day (Upper) Nest	0.0034	0.01364	0.0349	0.2855
B Carrier (Lower) Nest	0.0277	0.01736	0.1061	0.03978
L Carrier (Upper) Nest	0.0139	0.01714	0.0471	0.4955
L Time of Day (Lower) Nest	0.0144	0.007438	-0.0224	0.118
Allocation Parameters				
Phi Constant L Side	0.05	0.499	-0.8961	3.051
Phi Income (000) L Side	-0.00159	0.007447		
Phi Advance Purchase L Side	0.0185	0.03331		

Table 6.5. Summary of Various Models Estimated for Bimodal Dataset

		HeNGEV Model		NetGEV Model		NL (L) Model		NL (B) Model		MNL Model	
Log Likelihood at Convergence:		-174263.42		-174533.95		-174534.42		-174587.7		-178199.59	
Log Likelihood at Zeros:		-333220.45		-333220.45		-333220.45		-333220.45		-333220.45	
Rho Squared w.r.t. Zeros:		0.477		0.476		0.476		0.476		0.465	
	True Value	**Estimated Parameter**	**Std Err of Est**	**Estimated Parameter**	**Std Err of Est**	**Estimated Parameter**	**Std Err of Est**	**Estimated Parameter**	**Std Err of Est**	**Estimated Parameter**	**Std Err of Est**
Departure Time											
Before 08:00	0	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a
08:00-09:59	0.15	0.1775	0.0164	0.2122	0.0208	0.2263	0.0185	3.226	1.434	0.2687	0.0239
10:00-12:59	0.1	0.01935	0.1266	0.05807	0.1495	-0.03242	0.1313	-9.764	5.012	-5.258	0.3782
13:00-15:59	0.05	0.08199	0.02451	0.09039	0.09349	-0.007432	0.02562	-8.52	4.302	0.3407	0.02855
16:00-18:59	0.1	0.1058	0.01836	0.1213	0.06865	0.05001	0.01986	-6.12	3.132	-0.2751	0.02301
19:00 or later	-0.3	-0.3872	0.1264	-0.3594	0.1631	-0.4736	0.1312	-12.39	6.036	-5.763	0.3783
Level of Service											
Nonstop	0	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a
Single Connect	-2.3	-2.415	0.1287	-2.431	0.2356	-2.655	0.1331	-25.66	11.42	-7.907	0.3779
Double Connect	-5.8	-5.959	0.1588	-6.013	0.4957	-6.523	0.1603	-61.6	27.28	-12.8	0.3884
Flight Characteristics											
Distance Ratio	-0.01	-0.01159	0.00089	-0.01373	0.00121	-0.01477	0.00088	-0.19	0.08432	-0.07893	0.00136
Fare Ratio	-0.004	-0.004766	0.00045	-0.005724	0.00067	-0.006307	0.00048	-0.09279	0.04115	-0.0383	0.00046
Nesting Parameters											
B Time of Day (Upper) Nest	0.8	0.8034	0.01364	0.8349	0.2855			9.063	4.007		
B Carrier (Lower) Nest	0.2	0.2277	0.01736	0.3061	0.03978			3.163	1.4		
L Carrier (Upper) Nest	0.7	0.7139	0.01714	0.7471	0.4955	0.8746	0.01125				
L Time of Day (Lower) Nest	0.3	0.3144	0.00744	0.2776	0.118	0.3211	0.00668				
Allocation Parameters											
Phi Constant L Side	1	1.05	0.499	0.1039	3.051						
Phi Income (000) L Side	-0.04	-0.04159	0.00745								
Phi Advance Purchase L Side	0.2	0.2185	0.03331								

The use of the models for forecasting is evaluated by generating a secondary data set, using the same generation algorithm, and comparing the performance of the probabilities of the models (and their predicted market shares) against the generated choices. The relative performance of the HeNGEV model and the NetGEV model across the entire market are roughly similar, as can be seen in Table 6.6, and depicted in the right side of Figure 6.2. The two models over- or under-predict in roughly the same amounts for each itinerary. However, when the predictions are segmented by income as in Table 6.7, the HeNGEV model can be seen to outperform the NetGEV model in the extreme (high and low) income segments. Figure 6.3 highlights this improved performance, comparing the prediction errors in various income segments. In the top and bottom fifths of income, the HeNGEV model provides notably smaller errors in prediction than the NetGEV model. The NetGEV tends to predict larger but offsetting errors, with over prediction for the higher incomes and under prediction for lower incomes, or vice versa, thus creating similar performances at the market level, but inferior predictive performance in some income segments. This advantage for the HeNGEV model is not surprising, as income is not reflected in the NetGEV model at all. However, it is interesting the note that the benefits of the HeNGEV model are closely tied to the competitive dynamic of the alternatives. Alternatives 1 and 2 are the two available non-stop itineraries, but alternative 2 shows a much stronger income-related tilt in choice probabilities, resulting from a differing competitive environment for itineraries compared against the two non-stop time slots.

Table 6.6. HeNGEV and NetGEV Market-Level Predictions for Bimodal Dataset

	Total	Predictions		Differences	
Itinerary	**Observed**	**HeNGEV**	**NetGEV**	**HeNGEV**	**NetGEV**
1	44997	45280.64	45276.60	283.64	279.60
2	27580	27309.66	27311.60	-270.34	-268.40
3	2469	2559.70	2558.25	90.70	89.25
4	1418	1365.03	1358.40	-52.97	-59.60
5	1333	1365.03	1358.40	32.03	25.40
6	3395	3333.51	3330.60	-61.49	-64.40
7	1385	1365.03	1358.40	-19.97	-26.60
8	3073	2993.22	2995.20	-79.78	-77.80
9	2351	2459.39	2467.05	108.39	116.05
10	9	9.18	8.20	0.18	-0.80
11	4	4.00	2.95	0.00	-1.05
12	436	438.81	436.80	2.81	0.80
13	5	7.15	7.00	2.15	2.00
14	18	20.35	21.70	2.35	3.70
15	17	20.35	21.70	3.35	4.70
16	1160	1144.68	1158.45	-15.32	-1.55
17	3996	3929.32	3936.85	-66.68	-59.15
18	3294	3221.46	3226.20	-72.54	-67.80
19	2042	2161.66	2140.45	119.66	98.45
20	867	881.17	895.20	14.17	28.20
21	0	0.00	0.00	0.00	0.00
22	0	0.00	0.00	0.00	0.00
23	0	0.00	0.00	0.00	0.00
24	0	0.00	0.00	0.00	0.00
25	0	0.00	0.00	0.00	0.00
26	30	26.13	30.20	-3.87	0.20
27	55	52.26	49.85	-2.74	-5.15
28	66	52.26	49.85	-13.74	-16.15

Table 6.7. HeNGEV and NetGEV Predictions Segmented by Income for Bimodal Dataset

	Observed Choices					HeNGEV Model					NetGEV Model				
Itin	Bottom Fifth		Middle Fifth		Top Fifth	Bottom Fifth		Middle Fifth		Top Fifth	Bottom Fifth		Middle Fifth		Top Fifth
1	8866	8970	8992	9148	9021	100.3	47.6	62.0	-53.4	127.2	189.3	85.3	63.3	-92.7	34.3
2	5266	5342	5547	5593	5832	-64.9	8.0	-91.3	-19.3	-102.8	196.3	120.3	-84.7	-130.7	-369.7
3	536	505	511	463	454	25.5	28.2	2.1	27.7	7.2	-24.4	6.6	0.6	48.7	57.7
4	322	312	277	266	241	-13.3	-23.7	-3.1	-8.3	-4.5	-50.3	-40.3	-5.3	5.7	30.7
5	294	287	270	251	231	14.7	1.3	3.9	6.7	5.5	-22.3	-15.3	1.7	20.7	40.7
6	686	678	673	656	702	-12.0	-8.2	-6.1	7.6	-42.8	-19.9	-11.9	-6.9	10.1	-35.9
7	358	270	277	271	209	-49.3	18.3	-3.1	-13.3	27.5	-86.3	1.7	-5.3	0.7	62.7
8	683	645	625	571	549	-17.3	-17.6	-24.8	-1.1	-19.1	-84.0	-46.0	-26.0	28.0	50.0
9	509	511	447	452	432	17.2	-4.4	45.7	25.2	24.8	-15.6	-17.6	46.4	41.4	61.4
10	2	3	1	3	0	0.8	-0.8	0.9	-1.6	0.9	-0.4	-1.4	0.6	-1.4	1.6
11	1	1	0	1	1	0.3	0.0	0.8	-0.4	-0.7	-0.4	-0.4	0.6	-0.4	-0.4
12	74	72	90	104	96	4.1	11.6	-2.5	-12.1	1.7	13.4	15.4	-2.6	-16.6	-8.6
13	3	1	1	0	0	-0.8	0.8	0.5	1.1	0.6	-1.6	0.4	0.4	1.4	1.4
14	6	4	6	0	2	0.1	0.9	-1.9	3.2	0.0	-1.7	0.3	-1.7	4.3	2.3
15	3	6	2	4	2	3.1	-1.1	2.1	-0.8	0.0	1.3	-1.7	2.3	0.3	2.3
16	173	205	229	260	293	11.6	4.9	-1.1	-12.1	-18.6	58.7	26.7	2.7	-28.3	-61.3
17	837	838	811	763	747	-8.7	-33.9	-24.1	4.7	-4.6	-49.6	-50.6	-23.6	24.4	40.4
18	757	697	653	599	588	-34.1	-19.0	-6.8	11.6	-24.2	-111.8	-51.8	-7.8	46.2	57.2
19	423	466	397	397	359	44.9	-18.4	36.2	20.1	36.9	5.1	-37.9	31.1	31.1	69.1
20	158	149	165	170	225	-12.1	14.2	10.5	19.2	-17.7	21.0	30.0	14.0	9.0	-46.0
21	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
22	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
23	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
24	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
26	5	8	6	4	7	0.1	-2.8	-0.8	1.3	-1.6	1.0	-2.0	0.0	2.0	-1.0
27	14	15	9	13	4	0.0	-3.0	1.5	-4.1	2.8	-4.0	-5.0	1.0	-3.0	6.0
28	24	15	11	11	5	-10.0	-3.0	-0.5	-2.1	1.8	-14.0	-5.0	-1.0	-1.0	5.0
Total Absolute Deviation:						445	271.6	332.2	256.9	473.3	972.4	573.6	329.7	548.3	1046

Figure 6.2. Observations and Market-Level Prediction Errors, Bimodal Dataset

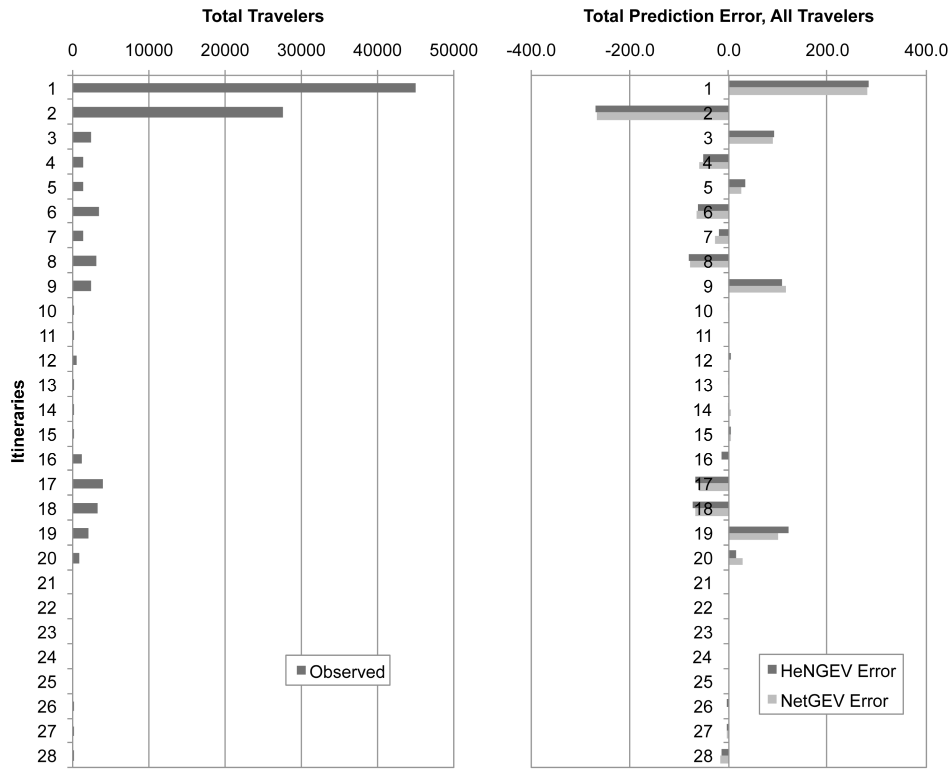
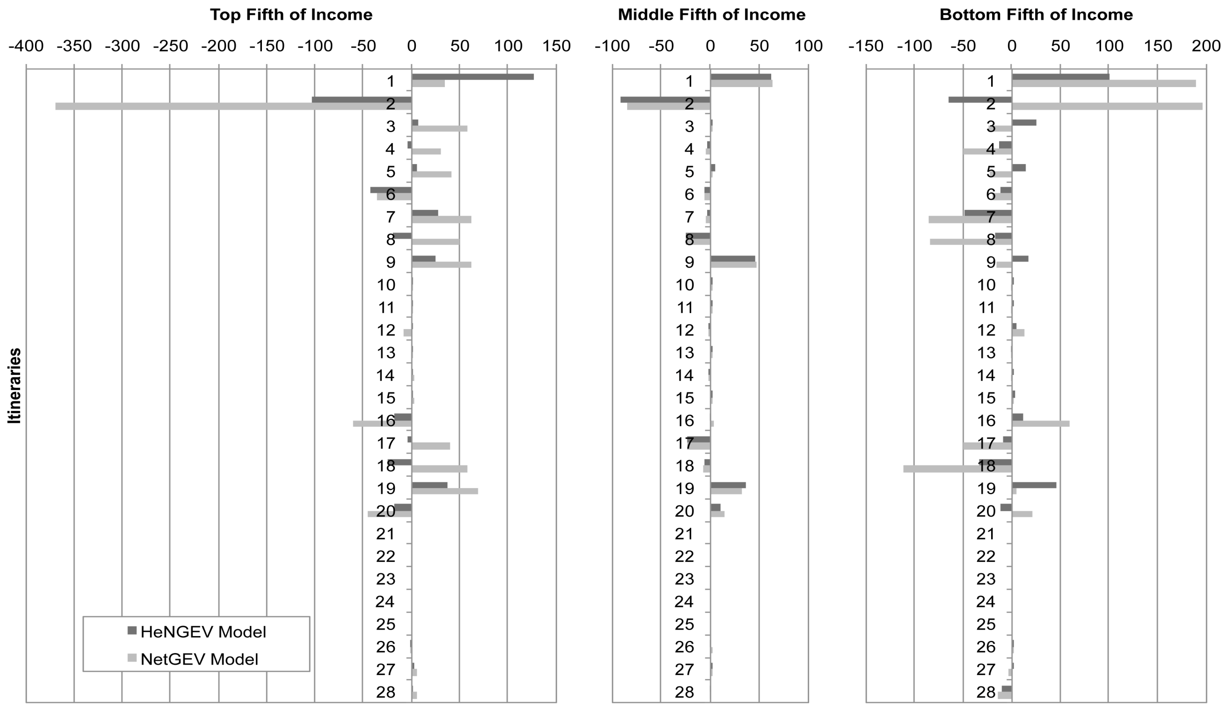


Figure 6.3. Prediction Errors, Segmented by Income, Bimodal Dataset



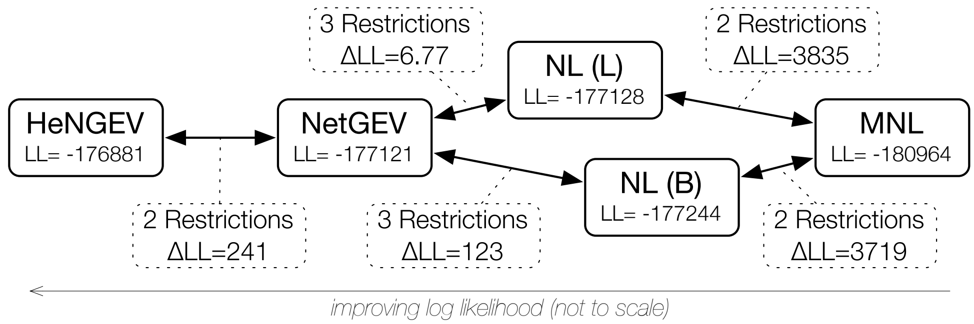
6.2.2. Unimodal Dataset

The estimated parameters for the HeNGEV model for the unimodal dataset are shown in Table 6.8. Most of the parameters in this model closely match the known true parameters, although three, with bolded t-statistics, show a statistically significant difference from the true values. That these three parameters are not correctly finding their true values is explained in part by the high correlation in their estimators, highlighted in Table 6.9.

As expected, the NetGEV model (shown in Table 6.10) performs relatively well on the unimodal dataset, as many more of the simulated travelers are concentrated in only one region of allocation values, instead of two. This concentration means that the NetGEV structure can more closely match the observed choice correlations without differentiating on the attributes of the travelers. Nevertheless, the NetGEV model still performs worse than the HeNGEV model. The NetGEV model has a log likelihood at convergence that is 240 smaller than the HeNGEV model, a highly significant deterioration given that only two degrees of freedom are lost, even in light of the large simulated dataset. The performance of the individual parameter estimates in the NetGEV and HeNGEV models are compared in Table 6.11. For each parameter in the model, the HeNGEV estimate is closer to the known true value than the NetGEV estimate, generally by about half. Further, the standard errors of the estimates are all smaller for the HeNGEV model, also by about half.

For a more complete picture, regular nested logit models were estimated using each of the two sub-models, as well as a multinomial logit model that ignored the error covariance entirely. The results of these models are shown in Table 6.12. A graphical representation of the relationship between the various estimated models for the unimodal data is shown in Figure 6.4. Not surprisingly, the MNL model with similarly defined utility functions performs relatively poorly, with log likelihood benefits in the thousands for a change to either nested structure.

Figure 6.4. Log Likelihoods and Relationships Between Models Estimated Using Unimodal Dataset



The L-only structure has a better fit for the data than the B-only model. This is consistent with the construction of this dataset, which is heavily weighted with decision makers exhibiting error correlation structures that are nearly the same as the L-only model. This heavy weight towards the L model is also reflected in the very small improvement (6.77) in log likelihood when moving from the L-only model to the NetGEV model, which incorporates both L and B submodels. While this change is still statistically significant (χ^2=13.54, with 3 degrees of freedom, p=0.0036) it is tiny compared to the changes observed between other models. In this instance, with most travelers exhibiting similar L choice patterns, it appears that upgrading to the NetGEV model alone does not provide much benefit. Far more improvement in the log likelihood is made when the heterogeneous covariance is introduced, which allows the small portion of the population that exhibits "B" choice patterns to follow that model, without adversely affecting the predictions for the larger L population.

Table 6.8. HeNGEV Model for Unimodal Dataset

Log Likelihood at Convergence:		-176880.64		
Log Likelihood at Zeros:		-333220.45		
Rho Squared w.r.t. Zeros:		0.469		
	True Value	**Estimated Parameter**	**Std. Error of Estimate**	**t Statistic vs. True**
Departure Time				
Before 08:00	0	0	n/a	n/a
08:00-09:59	0.15	0.1065	0.01796	**-2.42**
10:00-12:59	0.1	0.09257	0.09851	-0.08
13:00-15:59	0.05	0.02468	0.02453	-1.03
16:00-18:59	0.1	0.07013	0.01867	-1.60
19:00 or later	-0.3	-0.2975	0.09828	0.03
Level of Service				
Nonstop	0	0	n/a	n/a
Single Connect	-2.3	-2.286	0.1019	0.14
Double Connect	-5.8	-5.864	0.1354	-0.47
Flight Characteristics				
Distance Ratio	-0.01	-0.007141	0.001107	**2.58**
Fare Ratio	-0.004	-0.003359	0.0005518	1.16
Nesting Parameters				
B Time of Day (Upper) Nest	0.8	0.7994	0.01509	-0.04
B Carrier (Lower) Nest	0.2	0.1439	0.02585	**-2.17**
L Carrier (Upper) Nest	0.7	0.6746	0.01973	-1.29
L Time of Day (Lower) Nest	0.3	0.3075	0.006947	1.08
Allocation Parameters				
Phi Constant L Side	1	1.066	0.389	0.17
Phi Income (000) L Side	-0.03	-0.02912	0.005029	0.17
Phi Advance Purchase L Side	0.2	0.1772	0.02686	-0.85

Table 6.9. Parameter Estimator Correlation, HeNGEV Model for Unimodal Data

	08:00-09:59	10:00-12:59	13:00-15:59	16:00-18:59	19:00 or later	Distance Ratio	Fare Ratio	Single Connect	Double Connect	B Carrier (Lower) Nest	B Time of Day (Upper) Nest	L Time of Day (Lower) Nest	L Carrier (Upper) Nest	Phi Advance Purchase L Side	Phi Constant L Side	Phi Income (000) L Side
08:00-09:59	1.000	0.075	0.609	0.769	0.027	-0.901	-0.783	-0.124	-0.113	0.817	0.327	0.428	0.656	0.463	0.145	-0.411
10:00-12:59	0.075	1.000	0.052	0.132	0.996	-0.049	-0.026	0.958	0.737	0.061	-0.030	-0.317	0.059	0.022	0.004	-0.023
13:00-15:59	0.609	0.052	1.000	0.714	0.029	-0.547	-0.542	-0.050	0.006	0.561	-0.118	0.289	0.214	0.028	-0.075	-0.061
16:00-18:59	0.769	0.132	0.714	1.000	0.100	-0.661	-0.567	0.000	0.064	0.628	-0.016	0.216	0.354	0.132	-0.044	-0.150
19:00 or later	0.027	0.996	0.029	0.100	1.000	0.001	0.029	0.972	0.754	0.017	-0.070	-0.348	-0.007	-0.023	-0.011	0.016
Distance Ratio	-0.901	-0.049	-0.547	-0.661	0.001	1.000	0.870	0.133	0.141	-0.901	-0.336	-0.460	-0.685	-0.494	-0.168	0.439
Fare Ratio	-0.783	-0.026	-0.542	-0.567	0.029	0.870	1.000	0.198	0.220	-0.821	-0.409	-0.566	-0.723	-0.516	-0.155	0.461
Single Connect	-0.124	0.958	-0.050	0.000	0.972	0.133	0.198	1.000	0.800	-0.110	-0.230	-0.466	-0.185	-0.178	-0.075	0.146
Double Connect	-0.113	0.737	0.006	0.064	0.754	0.141	0.220	0.800	1.000	-0.112	-0.321	-0.419	-0.260	-0.265	-0.133	0.212
B Carrier (Lower) Nest	0.817	0.061	0.561	0.628	0.017	-0.901	-0.821	-0.110	-0.112	1.000	0.264	0.409	0.592	0.437	0.136	-0.398
B Time of Day (Upper) Nest	0.327	-0.030	-0.118	-0.016	-0.070	-0.336	-0.409	-0.230	-0.321	0.264	1.000	0.444	0.571	0.699	0.290	-0.598
L Time of Day (Lower) Nest	0.428	-0.317	0.289	0.216	-0.348	-0.460	-0.566	-0.466	-0.419	0.409	0.444	1.000	0.395	0.338	0.086	-0.304
L Carrier (Upper) Nest	0.656	0.059	0.214	0.354	-0.007	-0.685	-0.723	-0.185	-0.260	0.592	0.571	0.395	1.000	0.736	0.330	-0.598
Phi Advance Purchase L Side	0.463	0.022	0.028	0.132	-0.023	-0.494	-0.516	-0.178	-0.265	0.437	0.699	0.338	0.736	1.000	0.244	-0.702
Phi Constant L Side	0.145	0.004	-0.075	-0.044	-0.011	-0.168	-0.155	-0.075	-0.133	0.136	0.290	0.086	0.330	0.244	1.000	-0.811
Phi Income (000) L Side	-0.411	-0.023	-0.061	-0.150	0.016	0.439	0.461	0.146	0.212	-0.398	-0.598	-0.304	-0.598	-0.702	-0.811	1.000

Table 6.10. NetGEV Model for Unimodal Dataset

	Log Likelihood at Convergence:		-177121.27	
	Log Likelihood at Zeros:		-333220.45	
	Rho Squared w.r.t. Zeros:		0.468	
	True Value	**Estimated Parameter**	**Std. Error of Estimate**	**t Statistic vs. True**
Departure Time				
Before 08:00	0	0	n/a	n/a
08:00-09:59	0.15	0.06687	0.03759	**-2.21**
10:00-12:59	0.1	0.03704	0.1177	-0.53
13:00-15:59	0.05	-0.03495	0.07088	-1.20
16:00-18:59	0.1	0.02141	0.05334	-1.47
19:00 or later	-0.3	-0.3445	0.1120	-0.40
Level of Service				
Nonstop	0	0	n/a	n/a
Single Connect	-2.3	-2.331	0.1407	-0.22
Double Connect	-5.8	-5.956	0.2530	-0.62
Flight Characteristics				
Distance Ratio	-0.01	-0.004372	0.002449	**2.30**
Fare Ratio	-0.004	-0.002202	0.001068	1.68
Nesting Parameters				
B Time of Day (Upper) Nest	0.8	0.8307	0.1022	0.30
B Carrier (Lower) Nest	0.2	0.07244	0.04395	**-2.90**
L Carrier (Upper) Nest	0.7	0.6519	0.08702	-0.55
L Time of Day (Lower) Nest	0.3	0.3078	0.01321	0.59
Allocation Parameters				
Phi Constant L Side	1	0.5928	0.4722	-0.86

Table 6.11. Comparison of HeNGEV and NetGEV Models for Unimodal Data

	HeNGEV Model		NetGEV Model	
	Actual Error of Estimate	**Standard Error of Estimate**	**Actual Error of Estimate**	**Standard Error of Estimate**
Departure Time				
Before 08:00	n/a	n/a	n/a	n/a
08:00-09:59	-0.0435	0.01796	-0.08313	0.03759
10:00-12:59	-0.00743	0.09851	-0.06296	0.1177
13:00-15:59	-0.02532	0.02453	-0.08495	0.07088
16:00-18:59	-0.02987	0.01867	-0.07859	0.05334
19:00 or later	0.0025	0.09828	-0.0445	0.1120
Level of Service				
Nonstop	n/a	n/a	n/a	n/a
Single Connect	0.014	0.1019	-0.031	0.1407
Double Connect	-0.064	0.1354	-0.156	0.2530
Flight Characteristics				
Distance Ratio	0.002859	0.001107	0.005628	0.002449
Fare Ratio	0.000641	0.0005518	0.001798	0.001068
Nesting Parameters				
B Time of Day (Upper) Nest	-0.0006	0.01509	0.0307	0.1022
B Carrier (Lower) Nest	-0.0561	0.02585	-0.12756	0.04395
L Carrier (Upper) Nest	-0.0254	0.01973	-0.0481	0.08702
L Time of Day (Lower) Nest	0.0075	0.006947	0.0078	0.01321
Allocation Parameters				
Phi Constant L Side	0.066	0.389	-0.4072	0.4722
Phi Income (000) L Side	0.00088	0.005029		
Phi Advance Purchase L Side	-0.0228	0.02686		

Table 6.12. Summary of Various Models Estimated for Unimodal Dataset

		HeNGEV Model		NetGEV Model		NL (L) Model		NL (B) Model		MNL Model	
Log Likelihood at Convergence:		-176880.64		-177121.27		-177128.04		-177244.22		-180963.51	
Log Likelihood at Zeros:		-333220.45		-333220.45		-333220.45		-333220.45		-333220.45	
Rho Squared w.r.t. Zeros:		0.469		0.468		0.468		0.468		0.457	
	True Value	**Estimated Parameter**	**Std Err of Est**	**Estimated Parameter**	**Std Err of Est**	**Estimated Parameter**	**Std Err of Est**	**Estimated Parameter**	**Std Err of Est**	**Estimated Parameter**	**Std Err of Est**
Departure Time											
Before 08:00	0	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a
08:00-09:59	0.15	0.1065	0.01796	0.06687	0.03759	0.1615	0.01734	0.8323	0.1141	0.2668	0.02379
10:00-12:59	0.1	0.09257	0.09851	0.03704	0.1177	0.09445	0.1003	-1.326	0.4197	-4.684	0.2893
13:00-15:59	0.05	0.02468	0.02453	-0.03495	0.07088	-0.0211	0.02391	-1.303	0.3231	0.406	0.02834
16:00-18:59	0.1	0.07013	0.01867	0.02141	0.05334	0.04509	0.01896	-0.8219	0.2305	-0.1938	0.02282
19:00 or later	-0.3	-0.2975	0.09828	-0.3445	0.112	-0.3276	0.1001	-2.253	0.4913	-5.2	0.2894
Level of Service											
Nonstop	0	0	n/a	0	n/a	0	n/a	0	n/a	0	n/a
Single Connect	-2.3	-2.286	0.1019	-2.331	0.1407	-2.455	0.1019	-6.552	0.8812	-7.355	0.289
Double Connect	-5.8	-5.864	0.1354	-5.956	0.253	-6.274	0.1324	-16.19	2.098	-12.21	0.3015
Flight Characteristics											
Distance Ratio	-0.01	-0.007141	0.00111	-0.004372	0.00245	-0.01117	0.00081	-0.04809	0.00646	-0.07936	0.00136
Fare Ratio	-0.004	-0.003359	0.00055	-0.002202	0.00107	-0.005173	0.00045	-0.02619	0.00346	-0.03957	0.00046
Nesting Parameters											
B Time of Day (Upper) Nest	0.8	0.7994	0.01509	0.8307	0.1022			2.447	0.3128		
B Carrier (Lower) Nest	0.2	0.1439	0.02585	0.07244	0.04395			0.8607	0.111		
L Carrier (Upper) Nest	0.7	0.6746	0.01973	0.6519	0.08702	0.8193	0.01063				
L Time of Day (Lower) Nest	0.3	0.3075	0.00695	0.3078	0.01321	0.3133	0.0061				
Allocation Parameters											
Phi Constant L Side	1	1.066	0.389	0.5928	0.4722						
Phi Income (000) L Side	-0.03	-0.02912	0.00503								
Phi Advance Purchase L Side	0.2	0.1772	0.02686								

When applied for prediction, the models estimated from unimodal data perform in a similar manner as the bimodal models. The predictions of the HeNGEV model and the NetGEV model across the entire market are roughly similar, as can be seen in Table 6.13. The two models still over- or under-predict in roughly the same amounts for each itinerary. Again, when the predictions are segmented by income as in Table 6.14, the HeNGEV model can be seen to outperform the NetGEV model in all income segments, especially in the extremes of the income range. The errors for the whole market, on the right side of Figure 6.5, are roughly similar for both models. However, within the extreme high and low income segments (especially in the high income segment), as shown in Figure 6.6, the errors in prediction for the HeNGEV model are generally much smaller than those of the NetGEV model. These predictions show a similar pattern as observed in the bimodal dataset, with particularly large errors with offsetting signs appearing in the extreme income segments.

Table 6.13. HeNGEV and NetGEV Market-Level Predictions for Unimodal Dataset

	Total	Predictions		Differences	
Itinerary	**Observed**	**HeNGEV**	**NetGEV**	**HeNGEV**	**NetGEV**
1	45067	44806.47	44824.55	-260.53	-242.45
2	26746	26769.61	26753.7	23.61	7.70
3	2633	2649.82	2650.9	16.82	17.90
4	1346	1439.44	1432.45	93.44	86.45
5	1415	1439.44	1432.45	24.44	17.45
6	3521	3328.98	3355.5	-192.02	-165.50
7	1452	1439.44	1432.45	-12.56	-19.55
8	3328	3273.62	3293.55	-54.38	-34.45
9	2374	2485.81	2466.85	111.81	92.85
10	13	13.63	16.25	0.63	3.25
11	4	5.91	7.05	1.91	3.05
12	432	481.71	480.35	49.71	48.35
13	10	12.00	12	2.00	2.00
14	24	22.22	21.9	-1.78	-2.10
15	20	22.22	21.9	2.22	1.90
16	1047	1055.51	1053.15	8.51	6.15
17	3983	4014.62	4001.65	31.62	18.65
18	3412	3506.99	3506	94.99	94.00
19	2221	2257.96	2264.9	36.96	43.90
20	819	834.07	831.55	15.07	12.55
21	0	0.00	0	0.00	0.00
22	0	0.00	0	0.00	0.00
23	0	0.00	0	0.00	0.00
24	0	0.00	0	0.00	0.00
25	1	0.00	0	-1.00	-1.00
26	16	21.71	20.65	5.71	4.65
27	61	59.41	60.15	-1.59	-0.85
28	55	59.41	60.15	4.41	5.15

Table 6.14. HeNGEV and NetGEV Predictions Segmented by Income for Unimodal Dataset

Itin	Observed Choices: Bottom Fifth		Observed Choices: Middle Fifth		Observed Choices: Top Fifth	HeNGEV Model: Bottom Fifth		HeNGEV Model: Middle Fifth		HeNGEV Model: Top Fifth	NetGEV Model: Bottom Fifth		NetGEV Model: Middle Fifth		NetGEV Model: Top Fifth
1	8884	8958	9010	9139	9076	11.5	-27.3	-53.6	-152.0	-39.2	80.9	6.9	-45.1	-174.1	-111.1
2	5246	5211	5264	5423	5602	-128.2	33.0	72.5	23.3	23.0	104.7	139.7	86.7	-72.3	-251.3
3	572	565	533	500	463	-6.4	-18.5	-0.4	16.0	26.1	-41.8	-34.8	-2.8	30.2	67.2
4	275	285	280	277	229	48.0	19.2	10.5	-2.9	18.6	11.5	1.5	6.5	9.5	57.5
5	292	332	261	285	245	31.0	-27.8	29.5	-10.9	2.6	-5.5	-45.5	25.5	1.5	41.5
6	703	730	722	686	680	-37.0	-64.1	-56.2	-20.3	-14.4	-31.9	-58.9	-50.9	-14.9	-8.9
7	307	318	292	260	275	16.0	-13.8	-1.5	14.2	-27.4	-20.5	-31.5	-5.5	26.5	11.5
8	693	730	681	622	602	16.7	-49.7	-22.2	11.2	-10.4	-34.3	-71.3	-22.3	36.7	56.7
9	503	495	497	460	419	26.1	17.1	2.5	24.7	41.5	-9.6	-1.6	-3.6	33.4	74.4
10	6	3	0	1	3	-2.2	0.2	2.8	1.3	-1.6	-2.8	0.3	3.3	2.3	0.3
11	2	1	0	0	1	-0.3	0.4	1.2	1.0	-0.4	-0.6	0.4	1.4	1.4	0.4
12	78	78	84	95	97	12.7	15.7	11.9	3.6	5.8	18.1	18.1	12.1	1.1	-0.9
13	5	1	2	1	1	-1.6	1.9	0.5	1.0	0.3	-2.6	1.4	0.4	1.4	1.4
14	9	6	2	5	2	-2.8	-0.7	2.6	-1.3	0.4	-4.6	-1.6	2.4	-0.6	2.4
15	5	7	2	3	3	1.3	-1.7	2.6	0.7	-0.6	-0.6	-2.6	2.4	1.4	1.4
16	181	181	228	226	231	-11.2	10.9	-20.0	1.3	27.5	29.6	29.6	-17.4	-15.4	-20.4
17	842	803	822	761	755	-6.8	14.9	-16.7	29.3	10.9	-41.7	-2.7	-21.7	39.3	45.3
18	740	675	715	625	657	27.2	57.0	-8.7	50.7	-31.1	-38.8	26.2	-13.8	76.2	44.2
19	477	462	416	442	424	11.6	6.8	38.3	-4.9	-14.9	-24.0	-9.0	37.0	11.0	29.0
20	148	134	164	159	214	-7.3	20.7	0.9	18.0	-17.2	18.3	32.3	2.3	7.3	-47.7
21	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
22	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
23	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
24	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25	0	0	1	0	0	0.0	0.0	-1.0	0.0	0.0	0.0	0.0	-1.0	0.0	0.0
26	2	2	5	1	6	1.9	2.2	-0.7	3.5	-1.2	2.1	2.1	-0.9	3.1	-1.9
27	15	12	10	13	11	0.0	1.3	2.1	-2.3	-2.7	-3.0	0.0	2.0	-1.0	1.0
28	15	11	9	16	4	0.0	2.3	3.1	-5.3	4.3	-3.0	1.0	3.0	-4.0	8.0
Total Absolute Deviation:						407.9	407.2	361.9	399.5	321.9	530.6	519.2	369.9	564.4	884.2

Figure 6.5. Observations and Market-Level Prediction Errors, Unimodal Dataset

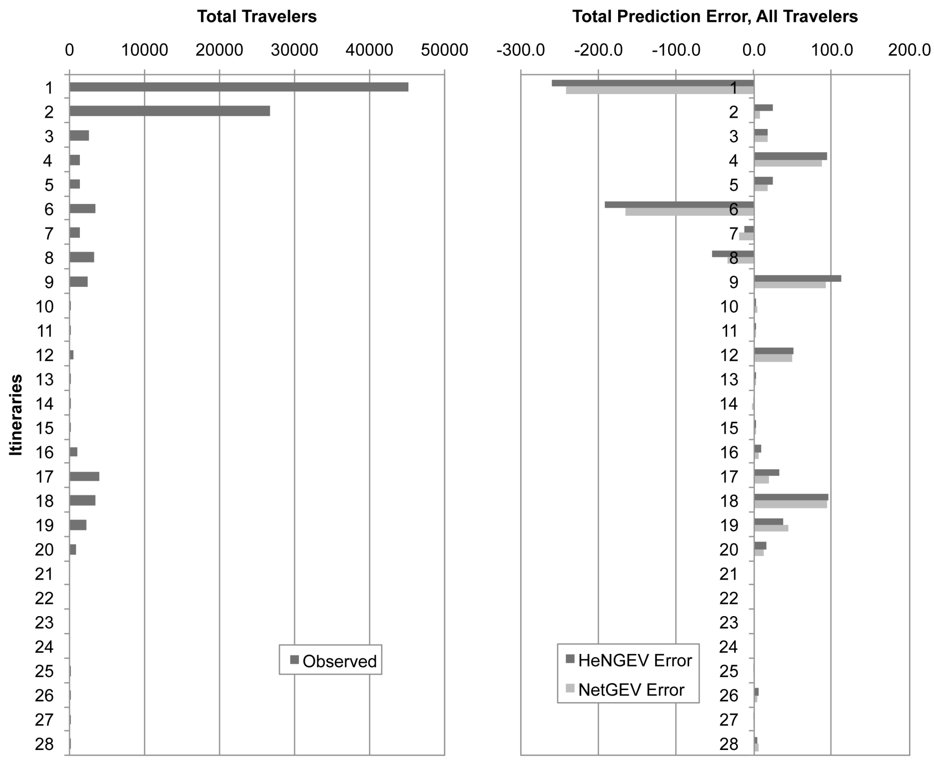
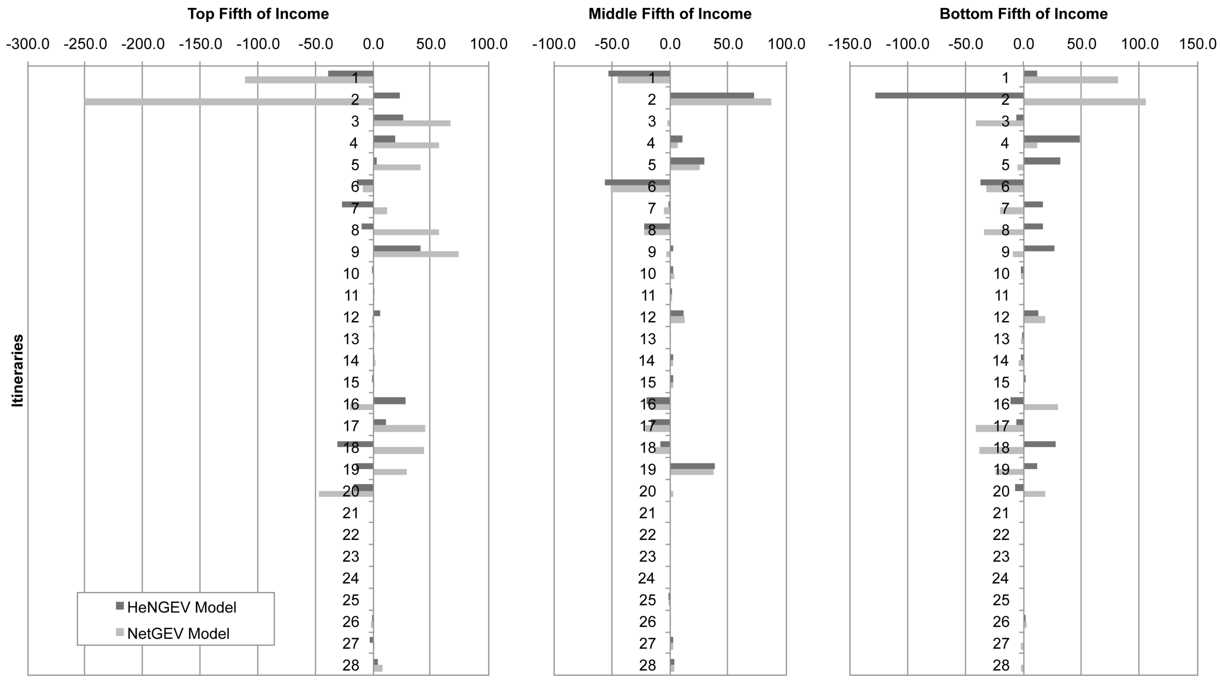


Figure 6.6. Prediction Errors, Segmented by Income, Unimodal Dataset



6.3. Discussion

Overall, the HeNGEV models show a better fit for the synthetic data than the matching homogeneous NetGEV models. The HeNGEV models give significantly better log likelihoods in both the bimodal and unimodal scenarios, indicating that this model type may be useful in a variety of situations, even when the fraction of the population exhibiting "unusual" behavior is small. Individual parameter estimates were generally improved by adopting the heterogeneous model, often by half or more of the error in the estimate.

Better fitting models are obviously an positive attribute of the HeNGEV structure, but they are not the only benefit. When used to predict choices of subsections of the population, the responsiveness of the correlation structure to data allows the HeNGEV to be a superior predictive tool. Such benefits could be especially appealing in revenue management systems, which seek specifically to segment markets in order to capture these types of differences in pricing and availability decisions.

CHAPTER 7

Summary and Conclusions

The research detailed in this dissertation refines and expands the Network GEV introduced by Daly and Bierlaire (2006). In particular, the allocation parameters of the NetGEV model are examined.

7.1. Normalization of Allocation Parameters

The allocation parameters of the NetGEV model must be normalized in order to be identified. Ideally, this normalization should be done in a manner that does not alter the relative expected values of the utilities of the alternatives, so that any preference or bias towards or against a particular alternative is expressed directly in the other model parameters. This dissertation introduces two algorithms for normalizing the allocation parameters in such non-biasing ways. The choice between these methods depends on the structure of the underlying network, with the crash-free method appropriate for networks that do not allow alternatives to correlate with themselves, and the crash-safe method appropriate for networks that allow independent adjustments to be applied to each alternative. When neither structural condition holds, a network transformation is shown to allow the crash-safe method to be implemented, although additional nonlinear constraints must be applied to the parameters.

It is generally preferable to avoid such constraints entirely. Therefore, a transformation of the allocation parameters is proposed, which relaxes even the linear constraints necessary for normalization. This transformation also allows the introduction of data into the allocation parameters, resulting in a error covariance matrix that is heterogeneous across decision makers.

7.2. Heterogeneous Network GEV Models

The resulting HeNGEV model offers several advantages over other advanced choice models, but suffers from some difficulties and limitations as well.

7.2.1. Advantages

Heterogeneous Covariance. The HeNGEV model can exhibit a heterogeneous covariance structure among alternatives, across decision makers. This is unusual in closed form models, but useful when identifiably different segments of the population are apparently differentially affected in their choices by unobserved attributes shared across sets of alternatives.

Closed Form. The HeNGEV model has a closed form probability expression, which does not require simulation to calculate. This is useful in both estimation and application of the models. In estimation, the [log] likelihoods must be calculated at each iteration in the non-linear optimization process. Since models require numerous such iterations to mind a maximum likelihood, the direct calculation of probabilities at each iteration accrues substantial benefits.

In application, the closed form of probabilities can also be an asset. Generating choices in a micro-simulation environment merely requires generating the alternative choice probabilities and making a single random draw to make a choice. The correlation structure evinced in the model is embedded in the probabilities, so that there is no need to generate multiple random variables, especially multiple correlated random variables. There is also no need (or opportunity) to use or save individual parameters, as would be typical in a mixed logit application. In very large simulations, this limited memory requirement can be particularly appealing.

7.2.2. Limitations

Network Specification. The nesting structure for HeNGEV models, while very flexible, still needs to be specified by the modeler. When there are a small number of possible structures that a modeler hypothesizes may be valid and active, it is possible to construct a model to estimate that contains them all. The resulting parameters of such a meta-model can be a guide to further refinements. However, when a modeler has no hypothesized structure to start with, the HeNGEV model itself cannot be used to build a generating network from scratch. While estimated parameters can be reviewed to find candidate nodes and arcs for removal, there is no output from the model to indicate where new nodes or arcs should be inserted to find a better model.

Parametric Explosion. There can quickly be a very large number of parameters. Within a HeNGEV model, there are separate parameters for each node, as well as potentially multiple separate parameters for each edge, all in addition to the parameters in the systematic utility. For most interesting networks, this will become an overwhelming number of parameters to estimate. Modelers will typically tackle this problem by grouping and restricting the values of the parameters, as seen in Chapter 6. However, the process for defining those grouping will require the exercise of modeler's judgement.

Conceptualization. Despite the obvious graphical representation of the underlying network, complex HeNGEV models can lack a simple conceptual form that modelers and policy makers can easily grasp. When limited to the more restrictive forms of the NL or GNL models, it is easy to translate the nesting structure into a conceptual form: nested alternatives are similar, deeper nests are more similar.

It seems natural to map the GEV generating network of a HeNGEV model into a choice process, with decision makers choosing an arc from each node progressively until reaching a final choice. Unfortunately, this leap is not an appropriate interpretation. The generating network represents a correlation structure among the random terms in the model, but the actual decision processes do not contain “error” in the same way. Decision makers are presumed to make choices in a deterministic manner, selecting from a choice set the single alternative with maximum utility. Instead of representing a choice process, the generating network is limited to expressing the relationships between the unobserved portions of the modeler’s representation of utility. For a single decision maker, all the various paths to each alternative in the GEV generating network exist and operate simultaneously.

For example, within the model explored in Chapter 6, the “B” and “L” sub-models do not directly correspond to a business traveler and a leisure traveler respectively. Instead, they represent the expected dominant portion of correlation for business and leisure travelers, but both categories draw somewhat from both models. A modeled business traveler would see same time, different carriers itineraries as more similar than same carrier, different time itineraries, but there would be less but still some correlation among the latter sets.

7.3. The Future

Much research remains to be done within discrete choice modeling, and with the HeNGEV model specifically. The obvious next steps are to explore the performance of HeNGEV models in real applications, as opposed to synthetic data sets. But other issues also remain for exploration in the theoretical understanding of the response in the networks.

Depending on the structure of the network, it is sometimes possible to create different models, with different joint probability density functions for the various error terms associated with the utilities of various alternatives, but that nevertheless have the same correlation matrix between those error terms. That is, the mapping from GEV generating networks to correlation matrices is a many-to-one map, instead of a one-to-one map. The joint probability density functions would still, however, differ in the higher moments. This implies that the mapping of correlation matrices to generating networks is a one-to-many map, i.e. there is not necessarily a single unique GEV generating network for any particular correlation matrix. This does not invalidate such models, but careful consideration might be required in selecting and using such models. More research is needed to fully understand the trade-offs between such allocative and inclusive correlation structures, and whether the differences would matter in any practical application.

Bibliography

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

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

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

Bhat, C. R., 1997a. Covariance heterogeneity in nested logit models: Econometric structure and application to intercity travel. Transportation Research Part B 31 (1), 11–21.

Bhat, C. R., 1997b. An endogeneous segmentation mode choice model with an application to intercity travel. Transportation Science 31 (1), 34–48.

Boxall, P. C., Adamowicz, W. L., 2002. Understanding heterogeneous preferences in random utility models: A latent class approach. Environmental and Resource Economics 23 (4), 421–446.

Bresnahan, T., Stern, S., Trajtenberg, M., 1997. Market segmentation and the sources of rents from innovation: personal computers in the late 1980s. RAND Journal of Economics 28 (0), S17–S44.

Chintagunta, P. K., Jain, D. C., Vilcassim, N. J., 1991. Investigating heterogeneity in brand preferences in logit models for panel data. Journal of Marketing Research 28 (4), 417–428.

Chu, C., 1989. Paired combinatorial logit model for travel demand analysis. In: Fifth World Conference on Transportation Research. Vol. 4. Ventura, CA, pp. 295–309.

Coldren, G. M., 2004. Modeling the competitive dynamic among air-travel itineraries with generalized extreme value models. Ph.D. thesis, Northwestern University.

Coldren, G. M., Koppelman, F. S., 2005. Modeling the competition among air-travel itinerary shares: GEV model development. Transportation Research Part A-Policy and Practice 39 (4), 345–365.

Daly, A., Bierlaire, M., May 2006. A general and operational representation of generalised extreme value models. Transportation Research Part B 40 (4), 285–305.

Germann, T. C., Kadau, K., Longini, I. M., Macken, C. M., 2006. Mitigation strategies for pandemic influenza in the united states. Proceedings of the National Academy of Sciences of the United States of America 103 (15), 5935–5940.

Kamakura, W. A., Russell, G. J., 1989. A probabilistic choice model for market segmentation and elasticity structure. Journal of Marketing Research 26 (4), 379–390.

Koppelman, F., Wen, C.-H., 1998. Alternative nested logit models: structure, properties and estimation. Transportation Research Part B 32 (5), 289–298.

Manski, C., 1977. The structure of random utility models. Theory and Decision 8 (3), 229–254.

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

McFadden, D., Train, K., 2000. Mixed mnl models for discrete response. Journal of Applied Econometrics 15, 447–470.

Planning Section MTC, March 2005. Report on findings of the peer review panel for the Metropolitan Transportation Commission (MTC).

Revelt, D., Train, K., 1998. Mixed logit with repeated choices: Households' choices of appliance efficiency level. Review of Economics and Statistics 80 (4).

Schrevens, E., Coppenolle, H., Portier, K. M., 2005. A comparative study between latent class binomial segmentation and mixed-effects logistic regression to explore between-respondent variability in visual preference for horticultural products. Journal of Applied Statistics 32 (6), 589–605.

Small, K. A., March 1987. A discrete choice model for ordered alternatives. Econometrica 55 (2), 409–424.

Swait, J., 1994. A structurreal equation model of latent segmentation and product choice for cross-sectional reveal preference choice data. Journal of Retailing and Consumer Services 1 (2), 77–89.

Swait, J., 2003. Flexible covariance structures for categorical dependent variables through finite mixtures of gev models. Journal of Business and Economics Statistics 21 (1), 80–87.

Train, K., 2003. Discrete choice methods with simulation. Cambridge University Press, New York.

Train, K. E., 1998. Recreation demand models with taste differences over people. Land Economics 74 (2), 230–239.

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

Waddell, P., Ulfarsson, G. F., Franklin, J. P., Lobb, J., 2007. Incorporating land use in metropolitan transportation planning. Transportation Research Part A: Policy and Practice 41 (5), 382–410.

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

APPENDIX A

Glossary of Mathematical Notation

A.1. Nodes and Choices

$\mathbf{i}, \mathbf{j}$	An elemental discrete choice
i, j, k	A node of the network, which can be associated with an elemental alternative, or any nest level
R	The root node of the network, i.e. the one node that has no predecessors
t	A decision maker

A.2. Edges and Paths

$\underset{\rightarrow}{ij}$	The edge connecting directly from i to j
$\vec{p}_{ij}$	A particular path connecting from i to j
$\mathrm{T}(k, j, i)$	The set of all paths from k to i that pass through j
$T(k, i)$	The set of all paths from k to i

A.3. Node Sets

$\mathcal{C}$	The set of nodes associated with elemental alternatives
$\mathcal{N}$	The set of network nodes excluding elemental alternatives
$\mathcal{Z}$	The set of all nodes in the network, $\mathcal{Z} = \mathcal{C} \cup \mathcal{N}$
$i^{\uparrow}$	The set of direct predecessor nodes to node i
$i^{\downarrow}$	The set of direct successor nodes from node i

$i^{\Downarrow}$ The set of direct and indirect (eventual) successor nodes from node i

$$i^{\Downarrow} = \begin{cases} i^{\downarrow} \cup \left\{ \bigcup_{j \in i^{\downarrow}} j^{\Downarrow} \right\} & i \in \mathcal{N} \\ \emptyset & i \in \mathcal{C} \end{cases}$$

$i^{\Uparrow}$ The set of direct and indirect (eventual) predecessor nodes from node i

$$i^{\Uparrow} = \begin{cases} i^{\uparrow} \cup \left\{ \bigcup_{j \in i^{\uparrow}} j^{\Uparrow} \right\} & i \neq R \\ \emptyset & i = R \end{cases}$$

A.4. Choice Indicators

δ^{*}_{ti} Indicator variable for t choosing node i specifically

$\delta^{\Downarrow}_{ti}$ Indicator variable for t choosing any sub-node of i

$$\delta^{\Downarrow}_{ti} = \sum_{j \in i^{\downarrow}} \delta_{tj}$$

δ_{ti} Indicator binary variable for t choosing node i or any sub-node of i

$$\delta_{ti} = \delta^{\Downarrow}_{ti} + \delta^{*}_{ti}$$

A.5. Parameters

μ_i Logsum parameter for network node i

β A vector of coefficients on attributes of alternatives, not associated with a node

ϕ_{ij} Vector of parameters corresponding to the edge connecting from i to j

Ξ A generic vector of all model parameters

$$\Xi \equiv \{\beta, \phi, \mu\}$$

A.6. Data (Independent Variables)

X_{ti} Vector of independent choice alternative attributes for alternative i of decision maker t; optionally (but often) includes alternative specific constants

Z_t Vector of independent allocation-related variables for decision maker t; optionally (but usually) includes a constant (i.e. $Z_t[0] = 1$)

A.7. Allocation

α_{tij} The edge allocative proportion corresponding to the connection from i to j for decision maker t

$\alpha_{tij} = \frac{\exp(\phi_{ij} Z_t)}{\sum\limits_{k \in j^{\uparrow}} \left[\exp\left(\phi_{kj} Z_t\right)\right]}$

$\vec{\alpha}_{p_{Ri}}$ The path allocative proportion corresponding to the path p from R to i

$\tilde{\alpha}_{Rji}$ The total of path allocative proportions corresponding to all paths from R to i which pass through j

$\tilde{\alpha}_{Rji} = \sum_{p \in T(R,j,i)} \vec{\alpha}_{p_{Ri}}$

A.8. Utility

V_{ti} The systematic utility of elemental alternative node i for decision maker t

$$V_{ti} = \begin{cases} \beta X_{ti} & \forall t, i \in \mathcal{C} \\ \mu_i \log \left[\sum_{j \in i^{\downarrow}} \left[\alpha_{tij}^{1/\mu_i} \exp\left[V_{tj}/\mu_i\right]\right]\right] & \forall t, i \in \mathcal{N} \end{cases}$$

G_{ti} The G factor of node i for t

$G_{ti} = \exp\left[V_{ti}\right]$

A.9. Probability

$P_{ti|k}$ Conditional probability that t chooses node i given having chosen $k \in i^{\uparrow}$

$$P_{ti|k} = \frac{\alpha_{tki}^{1/\mu_k} \exp[V_{ti}/\mu_k]}{\sum\limits_{j \in k^{\downarrow}} \left[\alpha_{tkj}^{1/\mu_k} \exp[V_{tj}/\mu_k]\right]} = \frac{\alpha_{tki}^{1/\mu_k} \exp[V_{ti}/\mu_k]}{\exp[V_{tk}/\mu_k]} = \frac{\exp\left[\left(V_{ti} + \log\left[\alpha_{tki}\right]\right)/\mu_k\right]}{\exp[V_{tk}/\mu_k]} = \left(\frac{\alpha_{tki} G_{ti}}{G_{tk}}\right)^{1/\mu_k}$$

$\vec{P}_{tik}$ Arc probability that t chooses node i after having chosen $k \in i^{\uparrow}$

$\vec{P}_{tik} = P_{tk} P_{ti|k}$

$P_{ti\vec{p}_{Ri}}$ Path probability that t chooses node i using path $\vec{p}_{Ri}$

P_{ti} Unconditional probability that t chooses node i

$$P_{ti} = \sum_{k \in i^{\uparrow}} P_{tk} P_{ti|k} = \sum_{k \in i^{\uparrow}} \vec{P}_{tik}$$