Systematically heterogeneous covariance in network GEV models

Jeffrey Newman [*]

May 11, 2009

[*]Transport and Mobility Laboratory, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland, jeffrey.newman@epfl.ch

Abstract

Mixed logit models can represent heterogeneity across individuals, in both observed and unobserved preferences, but require computationally expensive calculations to compute probabilities. A few methods for including error covariance heterogeneity in a closed form models have been proposed, and this paper adds to that collection, introducing a new form of a Network GEV model that sub-parameterizes the allocation values for the assignment of alternatives (and subnests) to nests. This change allows the incorporation of systematic (non-random) error covariance heterogeneity across individuals, while maintaining a closed form for the calculation of choice probabilities. Also explored is a latent class model of nested models, which can similarly express heterogeneity. The heterogeneous models are compared to a similar model with homogeneous covariance in a realistic scenario, and are shown to significantly outperform the homogeneous model, and the level of improvement is especially large in certain market segments. The results also suggest that the two heterogeneous models introduced herein may be functionally equivalent.

1   Introduction

A great deal of work had been put into creating discrete choice models that represent heterogeneity across decision makers, but this effort has mainly been focused on heterogeneous parameters incorporated into a mixed logit framework. In some ways, mixed logit models are very good for expressing such heterogeneity. The two basic formulations of such models, the random parameters logit and the error components logit, while formally equivalent, neatly express heterogeneity with respect to observed and unobserved attributes, respectively. Ultimately, a mixed logit model can approximate any random utility model arbitrarily closely (McFadden & Train, 2000), but it is not always desirable to use such a model. Mixed logit models can be difficult to specify correctly (Garrow, 2004), and when the mixing distributions are continuous, lack a closed form expression for choice probabilities, and thus require the use of simulation (Train, 2003). Simulation is becoming easier with increasing computational power, but still requires substantial extra computational and memory resources. For large models or large data sets, simulation can be burdensome.

The few functional forms for incorporating covariance heterogeneity into logit models that have been explored include a nested logit model with covariance heterogeneity, proposed by (Bhat, 1997a), and the Mixed Covariance GEV, explored by (Hess et al., 2005). This paper will introduce a new method of incorporating covariance heterogeneity, called the Heterogeneous Covariance Network GEV (HeNGEV) model, and examine a potential practical application of such a model, using a synthetically generated dataset.

1.1 COVNL Model

The possibility of directly expressing covariance heterogeneity across individuals was first explored by (Bhat, 1997a), who developed a nested logit model with covariance heterogeneity (COVNL model). The COVNL model expressed covariance heterogeneity as a deterministic function of some attributes of the decision maker. In a single level nested model, this took a relatively simple form, replacing the nested logit model's usual logsum parameter μ:

$$\mu_t = F\left(\alpha + \gamma z_t\right) \tag{1}$$

where z_n is a vector of attributes of the decision maker t, α and γ are scalar and vector parameters respectively, and F is a transformational function with the properties that $F\left(-\infty\right) = 0$, $F\left(+\infty\right) = 1$, and $f(x) = \partial F(x)/\partial x > 0$. This worked well for single level nesting structures, as restricting μ to the range between 0 and 1 is sufficient for a nested logit model to remain a GEV model consistent with utility maximization, per McFadden's (McFadden, 1978) original requirements.

When extending this structure to multiple levels, such that nest n is inside a higher level nest m, it becomes more complicated, as in addition to restricting the μ values to the (0,1] interval, they must also be restricted such that $\mu_n < \mu_m$. Bhat achieves this by defining the lower level logsum parameter μ_{nt} as in (1), and the higher level logsum parameter as

$$\mu_{mt} = F\left(\alpha + \gamma z_t + G\left(\delta + \eta w_t\right)\right) \tag{2}$$

with G as a strictly positive function, and w_t, δ, and η mirroring z_t, α, and γ. The positivity of G will ensure that the input value for F in (2)

is greater than that in (1), ensuring the correct ordering of the logsum parameter values, no matter the values of the parameters or the individual attributes. Since the logsum parameter is deterministic for each person, the COVNL does retain a simple closed form probability expression, making its implementation not computationally burdensome.

(Koppelman & Sethi, 2005) incorporated the COVNL form into a broader model structure they dubbed the Heterogeneous Generalized Nested Logit (HGNL) model. This model brought the heterogeneity proposed by Bhat into a Generalized Nested Logit structure, along with heteroscedasticity of error variance across individuals, as proposed by (Swait & Adamowicz, 2001).

1.2 Mixed Covariance GEV

More recently, (Hess et al., 2005) examined a more general form of covariance heterogeneity in the Mixed Covariance GEV model (MCGEV). In this model, the logsum parameters μ assume a random distribution across individuals, instead of adopting the variable but deterministic form proposed by (Bhat, 1997a). This change provides the benefit of greater generality and flexibility, but at the cost of greatly increased conceptual complexity, in order to maintain the decreasing values of the logsum parameters in the network. Additionally, if the distribution of the logsum parameters is continuous, the model loses the closed form for the probability calculations.

In the MCGEV model, the probability formulation for each alternative is conditioned on the distribution of the logsum parameter,

$$P_t(i) = \int_\mu P_t(i|\mu)f(\mu|\Omega)d\mu \tag{3}$$

The attributes of the model thus depend on the specific form chosen for the distribution of $f(\mu|\Omega)$. In order to maintain the necessary conditions for μ, the selected distribution must be bounded by 0 and 1. For multi-level nesting structures, the parameters also need to obey the proper ordering, as the μ for a nest for any individual must be less than the μ for higher level nests for that same individual. This requirement is necessary to ensure consistency with utility maximization. (Hess et al., 2005) address this issue

4

by redefining μ to be a function of the predecessor, so that

$$\mu_{lower} = \mu_{upper}\hat{\mu}_{lower} \qquad (4)$$

The hatted versions of μ for each nest then become the relevant parameters in estimation, and (3) is changed to be conditioned on $\hat{\mu}$ instead of μ. By using this product, imposing the condition that $\hat{\mu}_{lower}$ is constrained to be inside the (0,1] interval will automatically ensure that $\mu_{lower} < \mu_{upper}$. This replaces an inequality constraint relating the values of two parameters with an inequality constraint regarding one parameter and a constant, which is simpler to handle when estimating parameters. This becomes more complicated for cross-nested structures, as the adjustment in (4) must be made across all higher nests:

$$\mu_{lower} = \min_{u \in uppers} \{\mu_u\}\hat{\mu}_{lower}$$

2   Incorporating Heterogeneity Through Allocations

Each of these methods for incorporating heterogeneity into a logit models comes with substantial drawbacks. The COVNL model restricts the form of covariance to a strictly hierarchical form, as in a traditional nested logit model. The Mixed Covariance GEV relaxes that constraint, but can become difficult to operationalize, given the constraints required on the logsum parameters. The model forms introduced in this section attack the problem from a completely different angle.

2.1   Disaggregation of Allocation

As examined in the previous section, all known previous attempts to incorporate heterogeneity in error covariance have focused on sub-parameterizing the logsum scaling parameter. This direction of thinking is intuitive, as it is these parameters that induce covariance between error terms, and it is convenient to start with the simplest logit model containing covariance, which is the nested logit model.

However, a more complex but flexible model, such as the Network GEV model (NetGEV), introduced by (Daly & Bierlaire, 2006), provides alternative ways of generating error covariance heterogeneity. In particular, the NetGEV model includes allocation parameters on network links, which can be sub-parameterized instead of the logsum parameters, to achieve the same goal.

The NetGEV model is defined by a finite, directed, connected, circuit-free network. Such a network has a single source (root) node that represents the complete model, a sink node for each elemental alternative in the choice set, and some set of intermediate nesting nodes. Each node is associated with a G function, with G function at elemental alternative node a being

$$G^a(y) = y_a.$$

Each other node i has G function that is calculated based on the G values of that node's direct successors:

$$G^i(y) = \left(\sum_{j \in i^\downarrow} \left[\left(\alpha_{ij} G^j(y) \right)^{1/\mu_i} \right] \right)^{\mu_i}, \tag{5}$$

where $i^\downarrow$ is the set of successor nodes to i, α_{ij} is an allocation parameter associated with the edge from i to j in the network, and μ_i is a scaling parameter associated with node i. This model thus requires a logsum parameter on every node other than the elemental alternatives, as well as an allocation parameter on each edge in the network. For the ultimate model to be a valid GEV model, and consistent with random utility maximization according to McFadden's (McFadden, 1978) criteria, the allocation parameters must be non-negative, and the logsum parameters must obey sequential inequality constraints, so that the μ value for any node is smaller than or equal to the μ values for all of its network predecessors.

It is the sequential inequality constraint on the μ values that makes them undesirable to sub-parameterize. Since each μ must be less than the minimum of its predecessors, it is not easy to formulate a model that will allow μ to vary across individuals. The allocation parameters, on the other hand, do not have such a constraint.

2.1.1 Relaxing Allocation Parameter Constraints

The normalization of the NetGEV model does requires some type of constraint on the allocation parameters. As described in (Newman, 2008), the exact form of this constraint can help ensure an unbiased model, but it is often be convenient to have these allocation parameters sum to one, such that $\sum_{h \in i^{\uparrow}} \alpha_{hi} = 1$ (where $i^{\uparrow}$ is the set of predecessors of i). Imposing this restriction directly on estimated parameters is certainly possible. However, this restriction can be relaxed by transforming the parameters using the familiar logit structure:

$$\alpha_{hi} = \frac{\exp{(\phi_{hi})}}{\sum_{k \in i^{\uparrow}} [\exp{(\phi_{ki})}]}. \tag{6}$$

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 can be an advantage in parameter estimation, as nonlinear optimization algorithms are generally easier to implement when there are no (or fewer) constraints on the parameters.

2.1.2 Incorporating Data in the 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:

$$\alpha_{thi} = \frac{\exp{(\phi_{hi}^{\star} + \phi_{hi} Z_t)}}{\sum_{k \in i^{\uparrow}} [\exp{(\phi_{ki}^{\star} + \phi_{ki} Z_t)}]}, \tag{7}$$

where $\phi_{hi}^{\star}$ is the baseline parameter as in (6), Z_t is a vector of data specific to decision maker t, and ϕ_{hi} 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 (defining a "link-specific" constant), we can simplify (7) to[1]

$$\alpha_{thi} = \frac{\exp\left(\phi_{hi}Z_t\right)}{\sum\limits_{k \in i^\uparrow} \left[\exp\left(\phi_{ki}Z_t\right)\right]}. \tag{8}$$

Thus the G function for nesting nodes becomes

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

The ϕ parameters are all link specific parameters, analogous to alternative specific parameters in an MNL model. As usual for "alternative" specific constants and variables in logit models, one of the vectors ϕ_{ij} 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 not only in determining the systematic (observed) utility, but also in determining the correlation structure for random (unobserved) utility. This heterogeneous covariance network GEV (HeNGEV) model thus allows both the amount and the form of covariance to vary across decision makers.

For example, consider an air itinerary and fare class choice model, built on a network model. The network is bifurcated into two sub-structures, one with itinerary nested inside fare class, and the other with fare class nested inside itinerary. Each particular potential ticket choice is partly allocated to both substructures. The allocation parameters could then vary based on ticket purchase timing, with travelers who purchase well in advance tending to choose based on one substructure, while travelers who purchase closer to the date of travel shifting to the other structure.

---

[1]A similar mathematical form has been employed by (Sener et al., 2008) to model spatial correlation. Despite the superficial similarity of Sener's equation (4) and (8), they actually represent different relationships. In Sener, the data z is the spatially-generated relationships between alternatives, indexed over alternatives i and j, while in (8), the data Z is attributes of the decision maker, indexed over all decision makers t; along with the appropriate matching ϕ parameters in each model.

Since the form of (8) is by construction strictly positive, the HeNGEV model already meets one of the conditions of the NetGEV formulation, that α is positive. As long as the non-increasing μ parameters condition also holds, the HeNGEV model will be consistent with utility maximization.

2.2   Discrete Mixtures of Nested Logit Models

Another approach to expressing heterogeneity in covariance is through a latent class type formulation. In the traditional use of this model form, each decision maker is assumed to be a member of one of an unknown but assumed number of classes. Class membership is unobserved, but modeled in some way so that each decision maker t has a probability $Q_t(k)$ of belonging to class k. Conditional on membership in a particular class, the choice probabilities are calculated normally for the class' kernel choice model. The unconditional choice probabilities must then be calculated from the class membership probabilities and the various class kernel models:

$$P_t(i) = \sum_{k \in Classes} P_t(i|k) Q_t(k).$$

Latent class discrete choice models in transportation research have generally employed multinomial logit models as the class kernel model, with some or all of the MNL parameters allowed to vary across the classes, as in (Bhat, 1997b) and (Greene & Hensher, 2003). However, there is no particular reason why the kernel models could not be nested logit or NetGEV models, so as to mix together various covariance structures, as suggested by (Walker & Ben-Akiva, 2002). In this scenario, the various β parameters to the alternative utility functions should be assumed to be consistent across the classes, with only the nesting structure varying across the classes. This formulation would result in a mixture of covariance structures, controlled by the parameters of the class membership probability function. When that function contains as input data some attributes of the decision maker, the resulting total model will have a heterogeneous covariance structure. This discrete mixing is subtly different from that proposed by (Hess, 2005, Ch. 7). His mixing was done by varying the logsum parameter directly, while the mixing here envisions mixing entire nesting network structures as a whole, without manipulating the logsum parameters therein directly.

This latent class structure can be interpreted in two ways: as a latent but real separation of individuals into distinct choice classes, or as a variable (but deterministic) mixing of utility covariances within the population. The first interpretation is analogous to the traditional latent class interpretation, where the modeler cannot identify the specific structure to which any given individual ascribes her choice, so he represents this modeling uncertainty in a probabilistic mixing fashion. The second interpretation involves the probabilistic mixing of choice structures within the choice process modeled at the individual decision maker level, so that the individual decision makers in the model have complex internal choice processes.

In cross-sectional data, this distinction is not important, but in panel data, it can be relevant. Under the second interpretation, the relationship between the decision makers and discrete classes breaks down; individual decision makers in this model do not belong to a distinct class, but rather the probabilistic representation of the model of their choice process is itself being probabilistically determined. Under the first interpretation, they would retain their class membership and the associated choice model structure over time. Thus, the probabilities for their choices would need to be aggregated over all their choices before being averaged across possible classes:

$$P_t(i_1, i_2, i_3, ..., i_{N_t}) = \sum_{k \in Classes} \left[Q_t(k) \prod_{n=0}^{N_t} P_t(i_n|k) \right].$$ (9)

In second interpretation, such intermediate accumulation of probabilities across individuals for the panelization would not be appropriate. The operational difference between these approaches, while interesting, is beyond the scope of this paper and left for future research.

3   Application

The heterogeneous covariance models, by their nature, are most useful for analyzing complex decisions. Choices where decision makers only have a small handful of options don't offer lots of opportunity for complex correlation structures. In complex choices with large choice sets, the benefits of this flexible model can become more apparent. One typical such decision occurs in air travel booking, where travelers must choose between a variety

of itineraries when selecting an airline ticket. To demonstrate the model, we can examine a hypothetical choice scenario, and compare similar models with homogeneous and heterogeneous covariance forms.

3.1 Data Generation

This scenario involves synthetically created data that would approximate what might be observed for a flight itinerary choice between two medium sized airports in the United States. There are a variety of itinerary options (non-stop, single connect, and double connect flights on five different carriers). From each itinerary, various data attributes are provided, including departure time, level of service (non-stop, single connect, double connect), carrier, 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 1.

The advantage of the HeNGEV model described here 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 data set also includes data on the annual income level of each decision maker, as well as the number of days in advance that the ticket was purchased. These attributes were generated for 100,000 simulated travelers.

The structure of this model, based on one by (Coldren & Koppelman, 2005), is depicted in Figure 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 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.

Since the data in this example is synthetic, the true model underlying the observations is known. In particular, the distribution of the covariance structure in the population is known and defined to be heterogeneous. This distribution is shown in Figure 2. A large share of the population is grouped near the right side, having a covariance structure nearly entirely

Table 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

Figure 1: Flight Itinerary Choice Model for Synthetic Data
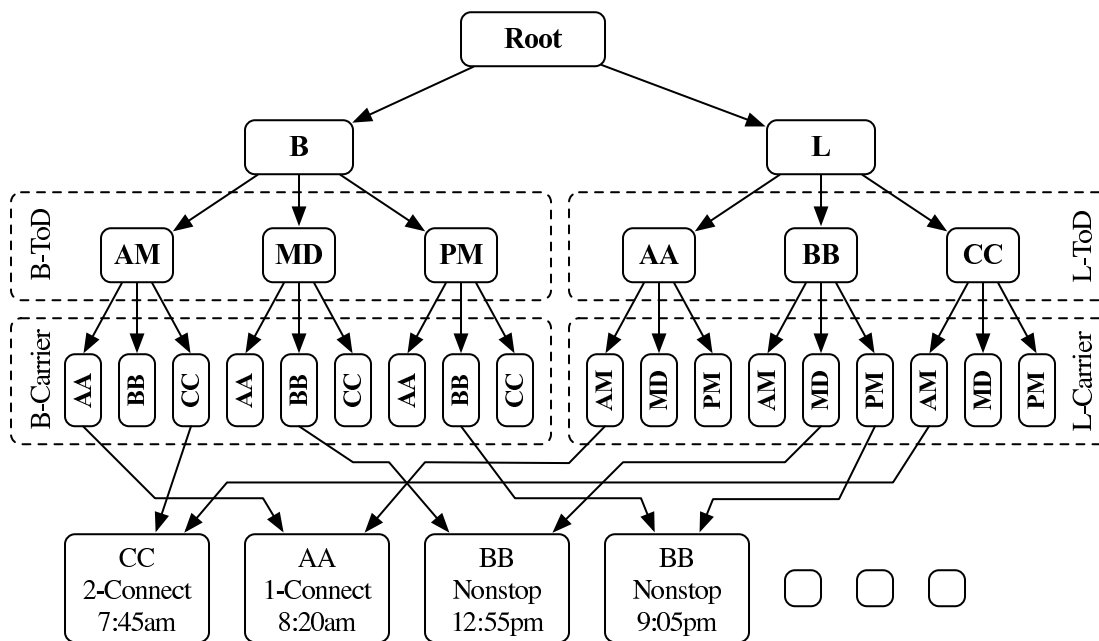
Adapted from (Coldren & Koppelman, 2005)

defined by the L submodel, while a much smaller share of the population is represented on the B submodel side. This reflects the common scenario in air travel, where there are a few (generally high-revenue and business-related) travelers, who make decisions in a different way than most other travelers, and would thus exhibit different error covariance structures.

3.2   Estimated Models

The estimated parameters for the HeNGEV model are shown in Table 2. 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 might be explained in part by the high correlation in their estimators, highlighted in Table 3. Such high correlations would tend to inflate the errors on the parameter's estimates (which should make the differences seem less significant, not more) so that these parameters having statistically significant errors is mildly troubling.
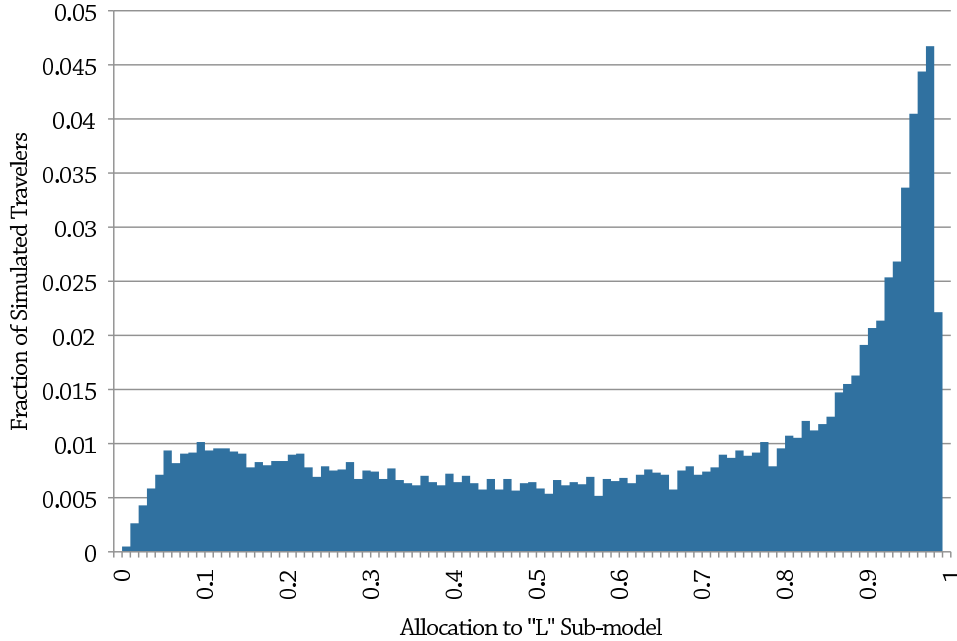
Table 2: HeNGEV Model

	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 3: Parameter Estimator Correlation, HeNGEV Model

	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	0.08	0.61	0.77	0.03	-0.90	-0.78	-0.12	-0.11	0.82	0.33	0.43	0.66	0.46	0.15	-0.41
10:00-12:59	0.08	1	0.05	0.13	1.00	-0.05	-0.03	0.96	0.74	0.06	-0.03	-0.32	0.06	0.02	0.00	-0.02
13:00-15:59	0.61	0.05	1	0.71	0.03	-0.55	-0.54	-0.05	0.01	0.56	-0.12	0.29	0.21	0.03	-0.08	-0.06
16:00-18:59	0.77	0.13	0.71	1	0.10	-0.66	-0.57	0.00	0.06	0.63	-0.02	0.22	0.35	0.13	-0.04	-0.15
19:00 or later	0.03	1.00	0.03	0.10	1	0.00	0.03	0.97	0.75	0.02	-0.07	-0.35	-0.01	-0.02	-0.01	0.02
Distance Ratio	-0.90	-0.05	-0.55	-0.66	0.00	1	0.87	0.13	0.14	-0.90	-0.34	-0.46	-0.69	-0.49	-0.17	0.44
Fare Ratio	-0.78	-0.03	-0.54	-0.57	0.03	0.87	1	0.20	0.22	-0.82	-0.41	-0.57	-0.72	-0.52	-0.16	0.46
Single Connect	-0.12	0.96	-0.05	0.00	0.97	0.13	0.20	1	0.80	-0.11	-0.23	-0.47	-0.19	-0.18	-0.08	0.15
Double Connect	-0.11	0.74	0.01	0.06	0.75	0.14	0.22	0.80	1	-0.11	-0.32	-0.42	-0.26	-0.27	-0.13	0.21
B Carrier (Lower) Nest	0.82	0.06	0.56	0.63	0.02	-0.90	-0.82	-0.11	-0.11	1	0.26	0.41	0.59	0.44	0.14	-0.40
B Time of Day (Upper) Nest	0.33	-0.03	-0.12	-0.02	-0.07	-0.34	-0.41	-0.23	-0.32	0.26	1	0.44	0.57	0.70	0.29	-0.60
L Time of Day (Lower) Nest	0.43	-0.32	0.29	0.22	-0.35	-0.46	-0.57	-0.47	-0.42	0.41	0.44	1	0.40	0.34	0.09	-0.30
L Carrier (Upper) Nest	0.66	0.06	0.21	0.35	-0.01	-0.69	-0.72	-0.19	-0.26	0.59	0.57	0.40	1	0.74	0.33	-0.60
Phi Advance Purchase L Side	0.46	0.02	0.03	0.13	-0.02	-0.49	-0.52	-0.18	-0.27	0.44	0.70	0.34	0.74	1	0.24	-0.70
Phi Constant L Side	0.15	0.00	-0.08	-0.04	-0.01	-0.17	-0.16	-0.08	-0.13	0.14	0.29	0.09	0.33	0.24	1	-0.81
Phi Income (000) L Side	-0.41	-0.02	-0.06	-0.15	0.02	0.44	0.46	0.15	0.21	-0.40	-0.60	-0.30	-0.60	-0.70	-0.81	1

Figure 2: Distribution of Allocation Weights in Unimodal Synthetic Data

Nevertheless, this model performs better than others on this dataset.

The NetGEV model without a heterogeneous covariance (shown in Table 4) performs relatively well, but definitely 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. The performance of the individual parameter estimates in the NetGEV and HeNGEV models are compared in Table 5. 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.

A latent class model was also estimated for the same dataset, with two classes representing the two correlation structures. The class membership model also utilized a multinomial logit structure, with class membership based on the same input variables as the allocation data in the HeNGEV model. The resulting parameter estimates and model fit were *identical* to those estimated using the HeNGEV model.

For a more complete picture, regular nested logit models were estimated

Table 4: NetGEV Model

	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 5: Comparison of HeNGEV and NetGEV Models

	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		

Figure 3: Log Likelihoods and Relationships Between Models Estimated Using Unimodal Dataset
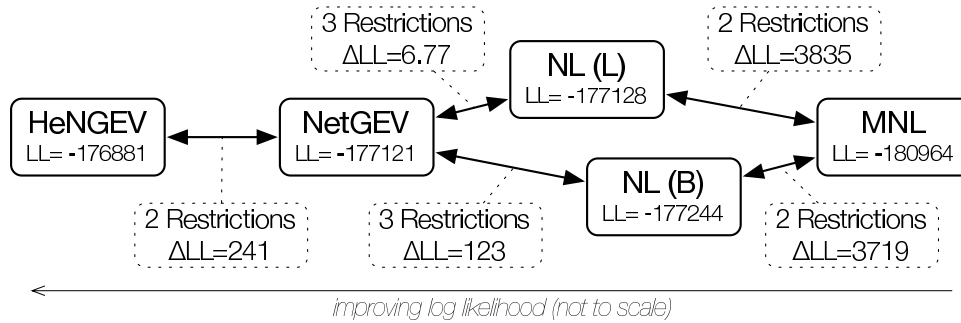
using each of the two sub-models, as well as a multinomial logit model that ignored the error covariance entirely. As shown in Figure 3, the results of these models were (by definition) inferior to even the NetGEV model, although the L-only nested submodel performed nearly as well as the homogeneous NetGEV model. 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. Of course, while all of the differences shown in Figure 3 are statistically significant, some differences are more significant than others.

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.

19

Table 6: HeNGEV and NetGEV Market-Level Predictions

		Predictions		Differences	
Itinerary	Total 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
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
Others	1455	1532.29	1531.95	77.29	76.95

When applied for prediction, the predictions of the HeNGEV model and the NetGEV model across the entire market are roughly similar, as can be seen in Table 6. 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 7, 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 4, 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 5, the errors in prediction for the HeNGEV model are generally much smaller than those of the NetGEV model. The overall market predictions for the NetGEV model end up close to the HeNGEV predictions because the particularly large errors appearing in the extreme income segments have offsetting signs.

Table 7: HeNGEV and NetGEV Predictions Segmented by Income

Itinerary	Observed Choices					HeNGEV Model					NetGEV Model				
	Bottom Fifth of Income	Lower Middle Income	Middle Fifth of Income	Upper Middle Income	Top Fifth of Income	Bottom Fifth of Income	Lower Middle Income	Middle Fifth of Income	Upper Middle Income	Top Fifth of Income	Bottom Fifth of Income	Lower Middle Income	Middle Fifth of Income	Upper Middle Income	Top Fifth of Income
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
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
Others	137	121	115	135	128	9	21.6	25.1	2.2	4.3	3	19.1	25.1	5.1	12.1
Total Absolute Deviation:						407.87	407.16	361.95	399.51	321.93	530.55	519.19	369.91	564.37	884.23

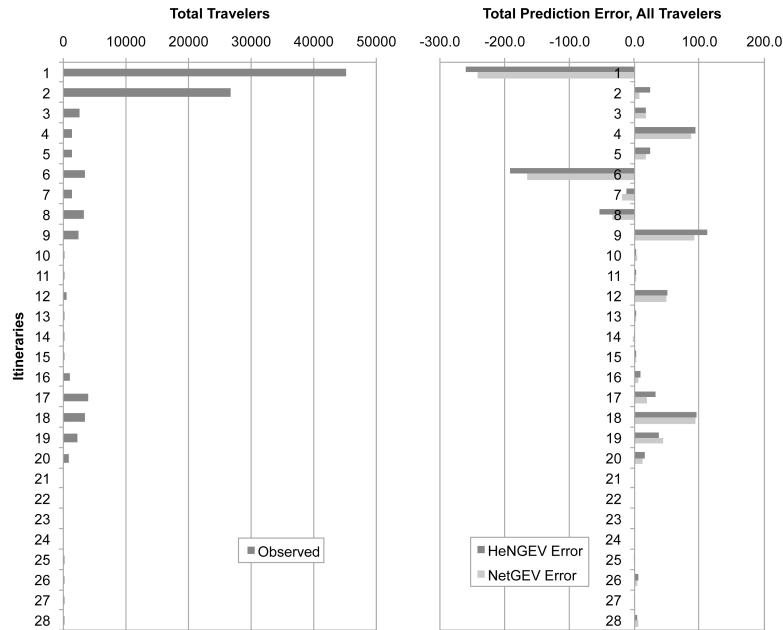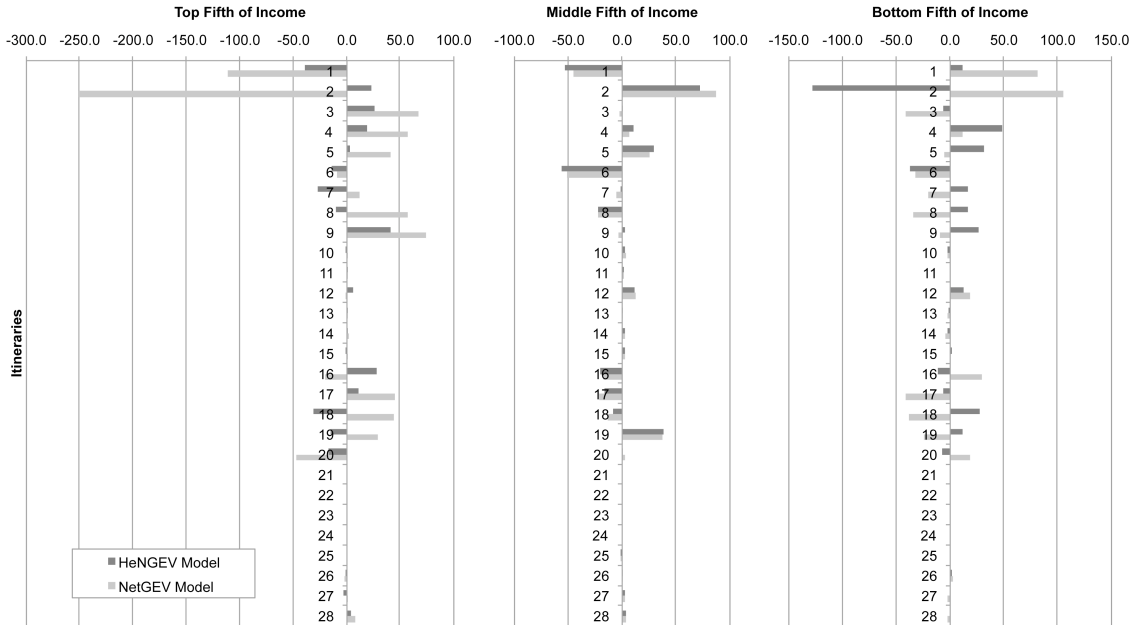Figure 4: Observations and Market-Level Prediction Errors

21

Figure 5: Prediction Errors, Segmented by Income

4 Conclusions

Overall, the HeNGEV models show a better fit for the synthetic data than the matching homogeneous NetGEV models. Individual parameter estimates were generally improved by adopting the heterogeneous model, often by half or more of the error in the estimate. This result is not especially surprising in this dataset, as it was constructed using a heterogeneous synthetic choice process. Nevertheless, the ability to recover the choice parameters, and to recover them more accurately than in a homogeneous model, demonstrates that the heterogeneous structure can be used successfully where appropriate.

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.

The identical results achieved for the HeNGEV and latent class models strongly suggests an equivalence between these two model forms. This may be a particular result of the covariance structure selected for this synthetic data experiment, although the topology of the network chosen (with all the allocation parameters at the bottom of the network, on edges connecting to the elemental alternatives) is reasonably general, as any NetGEV model can be transformed to have this bottom-only allocation (Newman, 2008). Additionally, the synthetic data was a cross-sectional dataset, and not panel data with identifiable repeated choices by a single decision maker. Nevertheless, the author conjectures that the network GEV model and the discrete mixture of nested logit model are indeed generally equivalent, although the full exploration of this is left for the future.

References

C. R. Bhat (1997a). 'Covariance Heterogeneity in Nested Models: Econometric Structure and Application to Intercity Travel'. *Transportation Research Part B: Methodological* **31**(1):11–21.

C. R. Bhat (1997b). 'An Endogeneous Segmentation Mode Choice Model with an Application to Intercity Travel'. *Transportation Science* **31**(1):34–48.

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

A. Daly & M. Bierlaire (2006). 'A general and operational representation of Generalised Extreme Value models'. *Transportation Research Part B: Methodological* **40**:285–305.

L. A. Garrow (2004). *Comparison of models representing correlation and random taste variation: An application to airline passengers' rescheduling behavior*. Ph.D. thesis, Northwestern University, Evanston, IL.

W. H. Greene & D. A. Hensher (2003). 'A latent class model for discrete choice analysis: contrasts with mixed logit'. *Transportation Research Part B: Methodological* **37**:681–698.

S. Hess (2005). *Advanced Discrete Choice Models with Applications to Transport Demand*. Ph.D. thesis, Imperial College London.

S. Hess, et al. (2005). 'Random Covariance Heterogeneity in Discrete Choice Models'. ERSA conference papers ersa05p375, European Regional Science Association.

F. S. Koppelman & V. Sethi (2005). 'Incorporating variance and covariance heterogeneity in the Generalized Nested Logit model: an application to modeling long distance travel choice behavior'. *Transportation Research Part B: Methodological* **39**(9):825 – 853.

D. McFadden (1978). 'Modeling the Choice of Residential Location'. In A. Karlqvist (ed.), *Spatial Interaction Theory and Residential Location*, pp. 75–96. North-Holland, Amsterdam.

D. McFadden & K. Train (2000). 'Mixed MNL Models For Discrete Response'. *Journal of Applied Econometrics* **15**:447–470.

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

I. N. Sener, et al. (2008). 'Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior'. Tech. rep., Department of Civil, Architectural & Environmental Engineering, The University of Texas at Austin, http://www.ce.utexas.edu/prof/bhat/ABSTRACTS/SpatialCorrelationPaper31July08.pdf.

J. Swait & W. L. Adamowicz (2001). 'Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice'. *Organizational Behavior and Human Decision Processes* **86**(2):141 – 167.

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

J. Walker & M. Ben-Akiva (2002). 'Generalized random utility model'. *Mathematical Social Sciences* **43**(3):303 − 343.