A Heteroscedastic Extreme Value Model of Intercity Mode Choice

Chandra R. Bhat

University of Massachusetts at Amherst

Abstract

Estimation of disaggregate mode choice models to estimate the ridership share on a proposed new (or improved) intercity travel service and to identify the modes from which existing intercity travelers will be diverted to the new or upgraded service constitutes a critical part of evaluating alternative travel service proposals to alleviate intercity travel congestion. This paper develops a new heteroscedastic extreme value model of intercity mode choice that overcomes the "independence of irrelevant alternatives" (IIA) property of the commonly used multinomial logit model. The proposed model allows a more flexible cross-elasticity structure among alternatives than the nested logit model. It is also simple, intuitive and much less of a computational burden than the multinomial probit model. The paper discusses the non-IIA property of the heteroscedastic extreme value model and presents an efficient and accurate Gaussian quadrature technique to estimate the heteroscedastic model using the maximum likelihood method. The multinomial logit, alternative nested logit structures, and the heteroscedastic model are estimated to examine the impact of improved rail service on business travel in the Toronto-Montreal corridor. The nested logit structures are either inconsistent with utility maximization principles or are not significantly better than the multinomial logit model. The heteroscedastic extreme value model, however, is found to be superior to the multinomial logit model. The heteroscedastic model predicts smaller increases in rail shares and smaller decreases in non-rail shares than the multinomial logit in response to rail-service improvements. It also suggests a larger percentage decrease in air share and a smaller percentage decrease in auto share than the multinomial logit. Thus, the multinomial logit model is likely to provide overly optimistic projections of rail ridership and revenue, and of alleviation in inter-city travel congestion in general, and highway traffic congestion in particular. These findings point to the limitations of the multinomial logit and nested logit models in studying intercity mode choice behavior and to the usefulness of the heteroscedastic model proposed in this paper.

1. Introduction

Increasing congestion on intercity highways and at intercity air terminals has raised serious concerns about the adverse impacts of such congestion on regional economic development, national productivity and competitiveness, and environmental quality. Recent studies (Transportation Research Board special report, 1991; Federal Aviation Administration report, 1987) suggest that intercity travel congestion is likely to grow even further through the next two decades. To alleviate such current and projected congestion, attention has been focused in recent years on identifying and evaluating alternative proposals to improve inter-city transportation services. Some of these proposals include construction of new (or expansion of existing) express roadways and airports (Moon, 1991), upgrading conventional rail services (KPMG Peat Marwick *et al.*, 1993), and construction of new high-speed ground transportation based on magnetic levitation technology (U.S. Army Corps of Engineers, 1990).

The large scale nature of the congestion alleviation proposals makes it imperative to undertake a careful *a priori* cost-benefit evaluation with respect to capital investment costs, environmental impacts, job market and economic development impacts, and revenues from the potential use of the new service. Among other things, such an evaluation entails the estimation of reliable intercity mode choice models to estimate ridership share on the proposed new (or improved) intercity service and to identify the modes from which existing intercity travelers will be diverted to the new (or upgraded) service. This paper develops a new heteroscedastic extreme value model of intercity mode choice that: (a) overcomes the "independence of irrelevant alternatives" (IIA) restriction of the commonly used multinomial logit model; (b) permits more flexibility in cross-elasticity structure than the nested logit model; and (c) is simple, intuitive, and computationally less burdensome compared to the multinomial probit model. The paper presents an efficient method to estimate the heteroscedastic extreme value model and compares the results obtained from applying the proposed model and the multinomial logit and nested logit models to the estimation of intercity travel mode choice in the Toronto-Montreal corridor.

The next section of the paper presents a background of intercity travel mode choice models and develops the motivation for the heteroscedastic extreme value model proposed in this paper. Section 3 advances the model structure for the heteroscedastic model. Section 4 discusses the non-IIA property of the model. Section 5 outlines the estimation procedure. Section 6 presents empirical results. The final section provides a summary of the research findings.

2. Intercity Travel Mode Choice Models: A Background

Intercity travel mode choice models are based on the utility maximization hypothesis which assumes that an individual's mode choice is a reflection of underlying preferences for each of the available alternatives and that the individual selects the alternative with the highest preference or utility. The utility that an individual associates with an alternative is specified to be the sum of a deterministic component (that depends on observed attributes of the alternative and the individual) and a random component (that represents the effects of unobserved attributes of the individual and unobserved characteristics of the alternative).

In most intercity mode choice models, the random components of the utilities of the different alternatives are assumed to be independent and identically distributed (IID) with a type I extreme value distribution (Johnson & Kotz, 1970, Chapter 21). This results in the multinomial logit model of mode choice (McFadden, 1973). The multinomial logit model has a simple and elegant closed-form mathematical structure, making it easy to estimate and interpret. However, it is saddled with the "independence of irrelevant alternatives" (IIA) property at the individual level (Ben-Akiva & Lerman, 1985); that is, the multinomial logit model imposes the restriction of equal cross-elasticities due to a change in an attribute affecting only the utility of an alternative i for all alternatives $j \neq i$. This property of equal proportionate change of unchanged modes is unlikely to represent actual choice behavior in many situations (Stopher *et al.*, 1981).

The rigid inter-alternative substitution pattern of the multinomial logit model can be relaxed by removing, fully or partially, the IID assumption on the random components of the utilities of the different alternatives. The IID assumption can be relaxed in one of three ways: (a) allowing the random components to be non-identical and non-independent (non-identical, non-independent random components); (b) allowing the random components to be correlated while maintaining the assumption that they are identically distributed (identical, but non-independent random components); and (c) allowing the random components to be non-identically distributed (different variances), but maintaining the independence assumption (non-identical, but independent random components). We briefly discuss each of these alternatives below.

Models with non-identical, non-independent random components commonly use a normal distribution for the error terms. The resulting model, referred to as the multinomial probit model, can accommodate a very general error structure. Unfortunately, the increase in flexibility of error structure comes at the expense of introducing several additional parameters in the covariance matrix. This generates a number of conceptual, statistical and practical problems, including difficulty in interpretation, highly non-intuitive model behavior, low precision of covariance parameter estimates, and increased difficulty in transferring models from one space-time sampling frame to another (see Horowitz, 1991; Currim, 1982). The multinomial probit choice probabilities also involve high dimensional integrals and this may pose computational problems when the number of alternatives exceeds four. The multinomial probit has rarely been used in travel demand modeling (for an application, see Bunch and Kitamura, 1990).

The distribution of the random components in models which use identical, non-independent random components is generally specified to be either normal or type I extreme value. Travel demand research has mostly used the type I extreme value distribution since it nests the multinomial logit. The resulting model, referred to as the nested logit model, allows partial relaxation of the assumption of independence among random components of alternatives (Daly & Zachary, 1979; McFadden, 1978). This model has a closed form solution, is relatively simple to estimate, and is more parsimonious than the multinomial probit model. However, it requires *a priori* specification of homogenous sets of alternatives for which the IIA property holds. This requirement has at least two drawbacks. First, the number of different structures to estimate in a search for the best structure increases rapidly as the number of alternatives increases. Second, the actual competition structure among alternatives may be a continuum which cannot be accurately represented by partitioning the alternatives into mutually exclusive subsets. The nested logit model has seldom been used in intercity mode choice modeling (see Forinash & Koppelman, 1993 for a recent application).

The concept that heteroscedasticity in alternative error terms (*i.e.*, independent, but not identically distributed error terms) relaxes the IIA assumption is not new (see Daganzo, 1979), but has received little (if any) attention in travel demand modeling and other fields. In fact, the IIA property has become virtually synonymous with the assumption of lack of similarity (or independence of random components) among the choice alternatives in travel demand literature. In his study, Daganzo (1979) used independent negative exponential distributions with different variances for the random error components to develop a closed-form discrete choice model which does not have the IIA property. However, his model has not seen much application since it requires that the perceived utility of any alternative not exceed an upper bound. Daganzo's model also does not nest the multinomial logit model.

The model developed in this paper falls under the final category of non-IID models. Specifically, we develop a random utility model with independent, but non-identical error terms distributed with a type I extreme value distribution.[1] This heteroscedastic extreme value model allows the utility of alternatives to differ in the amount of stochasticity (*i.e.*, allows different variances on the random components across alternatives). Unequal variances of the random components are likely to occur when the variance of an unobserved variable that affects choice is different for different alternatives. For example, in an intercity mode choice model, if comfort is an unobserved variable whose values vary considerably for the train mode (based on, say, the degree of crowding on different train routes) but little for the automobile mode, then the random components for the automobile and train modes will have different variances (Horowitz, 1981).

The heteroscedastic extreme value model developed here nests the restrictive multinomial logit model and is flexible enough to allow differential cross-elasticities among all pairs of alternatives. It does not require *a priori* identification of mutually exclusive market partitions as does the nested logit structure. It is more efficient in model structure specification than the nested logit formulation since a single model structure is to be estimated rather than testing different nested structures. On the other hand, it is parsimonious compared to the multinomial probit model introducing only J-1 additional parameters in the covariance matrix as opposed to [J*(J-1)/2]-1 additional parameters in the probit model (J is the total number of alternatives in the universal choice set). It also poses much less of a computational burden requiring only the evaluation of a 1-dimensional integral (independent of the number of alternatives) compared to the evaluation of a J-1 dimensional integral in the multinomial probit model. Finally, unlike the multinomial probit model, the heteroscedastic extreme value model is easy to interpret and its behavior is intuitive (as we discuss in Section 4 of the paper).

---

[1] The author recently became aware of a similar model proposed by Allenby and Ginter (1993) in a marketing context. However, the discussion of the properties of the model and the procedure to estimate the model are very different in the two research efforts.

3. Model Structure

The random utility of alternative i, U_i, for an individual in random utility models takes the form (we develop the model structure at the individual level and so do not use an index for individuals in the following presentation):

$$U_i = V_i + \varepsilon_i \tag{1}$$

where V_i is the systematic component of the utility of alternative i which is a function of observed attributes of alternative i and observed characteristics of the individual, and ε_i is the random component of the utility function. Let C be the set of alternatives available to the individual. We assume that the random components in the utilities of the different alternatives have a type I extreme value distribution and are independent, but non-identically distributed. We also assume that the random components have a location parameter equal to zero and a scale parameter equal to θ_i for the ith alternative.[2] Thus, the probability density function and the cumulative distribution function of the random error term for the ith alternative are:

$$f(\varepsilon_i) = \frac{1}{\theta_i} e^{-\frac{\varepsilon_i}{\theta_i}} e^{-e^{-\frac{\varepsilon_i}{\theta_i}}} \text{ and } F_i(z) = \int_{\varepsilon_i=\infty}^{\varepsilon_i=z} f(\varepsilon_i) d\varepsilon_i = e^{-e^{-\frac{z}{\theta_i}}}. \tag{2}$$

The random utility formulation of equation (1), combined with the assumed probability distribution for the random components in equation (2) and the assumed independence among the random components of the different alternatives, enables us to develop the probability that an individual will choose alternative $i (P_i)$ from the set C of available alternatives:

$$\begin{aligned} P_i &= \text{Prob}(U_i > U_j), \qquad \text{for all } j \neq i, j \in C \\ &= \text{Prob}(\varepsilon_j \leq V_i - V_j + \varepsilon_i), \quad \text{for all } j \neq i, j \in C \\ &= \int_{\varepsilon_i=-\infty}^{\varepsilon_i=+\infty} \prod_{j \in C,\, j \neq i} \Lambda\left[\frac{V_i - V_j + \varepsilon_i}{\theta_j}\right] \frac{1}{\theta_i} \lambda\left(\frac{\varepsilon_i}{\theta_i}\right) d\varepsilon_i \end{aligned} \tag{3}$$

---

[2] The assumption of a location parameter equal to zero for the random components is in no way restrictive as long as constants are included in the systematic utility for each alternative. Also note that the variance of the ith alternative's error term is $\pi^2\theta_i^2/6$.

where $\lambda(.)$ and $\Lambda(.)$ are the probability density function and cumulative distribution function of the standard type I extreme value distribution, respectively, and are given by (see Johnson & Kotz, 1970)

$$\lambda(t) = e^{-t}e^{-e^{-t}} \text{ and } \Lambda(t) = e^{-e^{-t}}. \tag{4}$$

Substituting $w = \varepsilon_i / \theta_i$ in equation (4), the probability of choosing alternative *i* can be re-written as follows

$$P_i = \int_{w=-\infty}^{w=+\infty} \prod_{j \in C, j \neq i} \Lambda\left[\frac{V_i - V_j + \theta_i w}{\theta_j} \right] \lambda(w) dw. \tag{5}$$

It can be proved that the probabilities given by the expression in equation (5) sum to one over all alternatives (see Appendix A for a proof). If the scale parameters of the random components of all alternatives are equal, then the probability expression in equation (5) collapses to that of the multinomial logit (see McFadden, 1973).

4. Non-IIA Property of the Heteroscedastic Extreme Value Model

The heteroscedastic extreme value model (or simply the heteroscedastic model) discussed in the previous section avoids the pitfalls of the IIA property of the multinomial logit model by allowing different scale parameters across alternatives. Intuitively, we can explain this by realizing that the error term represents unobserved characteristics of an alternative; that is, it represents uncertainty associated with the expected utility (or the systematic part of utility) of an alternative. The scale parameter of the error term, therefore, represents the level of uncertainty. It sets the relative weights of the systematic and uncertain components in estimating the choice probability. When the systematic utility of some alternative *l* changes, this affects the systematic utility differential between another alternative *i* and the alternative *l*. However, this change in the systematic utility differential is tempered by the unobserved random component of alternative *i*. The larger the scale parameter (or equivalently, the variance) of the random error component for alternative *i*, the more tempered is the effect of the change in the systematic utility differential (see the numerator of the cumulative distribution function term in equation 5) and smaller is the elasticity effect on the probability of choosing alternative *i*. In particular, two alternatives will have the same elasticity effect due to a change in the systematic utility of another alternative only if they have the same scale parameter on the random components. This property is a logical and intuitive extension of the case of the multinomial logit in which all scale parameters are constrained to be equal and, therefore, all cross-elasticities are equal.

Formally, the effect of a small change in the systematic utility of an alternative *l* on the probability of choosing alternative *i* may be written as:

$$\frac{\partial P_i}{\partial V_l} = \int_{w=-\infty}^{w=+\infty} -\frac{1}{\theta_l} \exp\left[\frac{-V_i + V_l - \theta_i w}{\theta_l}\right] \prod_{j \in C, j \neq i} \Lambda \left[\frac{V_i - V_j + \theta_i w}{\theta_j}\right] \lambda(w) dw \tag{6}$$

and the effect of a change in the systematic utility of alternative *i* on the probability of choosing *i* as:

$$\frac{\partial P_i}{\partial V_i} = -\sum_{l \in C, l \neq i} \frac{\partial P_i}{\partial V_l}. \tag{7}$$

Assuming a linear-in-parameters functional form for the systematic component of utility for all alternatives, the cross-elasticity for alternative *i* with respect to a change in the *k*th level of service variable in the *l*th alternative's systematic utility, x_{kl}, can be obtained as:

$$\eta_{x_{kl}}^{P_i} = \left[\frac{\partial P_i}{\partial V_l} / P_i\right] * \beta_k * x_{kl} \tag{8}$$

where β_k is the estimated coefficient on the level of service variable *k* (assumed to be generic across alternatives here). The corresponding self-elasticity for alternative *i* with respect to a change in x_{kl} is

$$\eta_{x_{kl}}^{P_i} = \left[\frac{\partial P_i}{\partial V_i} / P_i\right] * \beta_k * x_{ki} \tag{9}$$

The equivalence of the heteroscedastic model elasticities when all the scale parameters are identically equal to one and those of the multinomial logit model is straightforward to establish (the proof is available upon request from the author). If, however, the scale parameters are unconstrained as in the heteroscedastic model, then the relative magnitudes of the cross-elasticities of any two alternatives i and j with respect to a change in the level of service of another alternative l are characterized by the scale parameter of the random components of alternatives i and j

$$\eta_{x_{kl}}^{P_i} > \eta_{x_{kl}}^{P_j} \text{ if } \theta_i < \theta_j \text{ ; } \eta_{x_{kl}}^{P_i} = \eta_{x_{kl}}^{P_j} \text{ if } \theta_i = \theta_j \text{ ; } \eta_{x_{kl}}^{P_i} < \eta_{x_{kl}}^{P_j} \text{ if } \theta_i > \theta_j \text{ .} \tag{10}$$

This important property of the heteroscedastic model allows for a simple and intuitive interpretation of the model, unlike the multinomial probit where there is no easy correspondence between the covariance matrix of the random components and elasticity effects. One has to numerically compute the elasticities by evaluating multivariate normal integrals in the multinomial probit model to identify the relative magnitudes of cross-elasticity effects.

5. Model Estimation

The heteroscedastic extreme value model developed in this paper is estimated using the maximum likelihood technique. We assume a linear-in-parameters specification for the systematic utility of each alternative given by $V_{qi} = \beta' X_{qi}$ for the qth individual and ith alternative (we introduce the index for individuals in the following presentation since the purpose of the estimation is to obtain the model parameters by maximizing the likelihood function over all individuals in the sample). The parameters to be estimated in the heteroscedastic model are the parameter vector β and the scale parameters of the random component of each of the alternatives (one of the scale parameters is normalized to one for identifiability). The log likelihood function to be maximized can be written as

$$\text{‹} = \sum_{q=1}^{q=Q} \sum_{i \in C_q} y_{qi} \log \left\{ \int_{w=-\infty}^{w=+\infty} \prod_{j \in C_q,\, j \neq i} \Lambda \left[\frac{V_{qi} - V_{qj} + \theta_i w}{\theta_j} \right] \lambda(w) dw \right\}, \tag{11}$$

where C_q is the choice set of alternatives available to the qth individual and y_{qi} is defined as follows

$$y_{qi} = \begin{cases} 1 \text{ if the } q \text{ th individual chooses alternative } i \\ (q = 1, 2, \ldots, Q,\ i = 1, 2, \ldots, I\ ) \\ 0 \text{ otherwise.} \end{cases} \tag{12}$$

The log likelihood function in equation (11) has no closed-form expression. An improper integral needs to be computed for each alternative-individual combination at each iteration of the maximization of the log-likelihood function. The use of conventional numerical integration techniques (such as Simpson's method or Romberg integration) for the evaluation of such integrals is cumbersome, expensive and often leads to unstable estimates because they require the evaluation of the integrand at a large number of equally spaced intervals in the real line (Butler & Moffitt, 1982; Chintagunta *et al.*, 1991). On the other hand, Gaussian quadrature (Press *et al.*, 1986) is a more sophisticated procedure that can obtain highly accurate estimates of the integrals in the likelihood function by evaluating the integrand at a relatively small number of support points, thus achieving gains in computational efficiency of several orders of magnitude. However, to apply Gaussian quadrature methods, equation (11) must be expressed in a form suitable for application of one of several standard Gaussian formulas (see Stroud & Secrest, 1966 for a review of Gaussian formulas). To do so, define a variable $u = e^{-w}$. Then, $\lambda(w)dw = -e^{-u}du$ and $w = -\ln u$. Also define a function G_{qi} as

$$G_{qi}(u) = \prod_{j \in C_q, j \neq i} \Lambda\left[\frac{V_{qi} - V_{qj} - \theta_i \ln u}{\theta_{\mathrm{j}}}\right] \tag{13}$$

Then we can re-write equation (11) as

$$\text{‹} = \sum_q \sum_{i \in C_q} y_{qi} \log\left\{\int_{u=0}^{u=\infty} G_{qi}(u) e^{-u} du\right\} \tag{14}$$

The expression within braces in the above equation can be estimated using the Laguerre Gaussian quadrature formula, which replaces the integral by a summation of terms over a certain number (say *K*) of support points, each term comprising the evaluation of the function $G_{qi}(.)$ at the support point *k* multiplied by a probability mass or weight associated with the support point (the support points are the roots of the Laguerre polynomial of order *K* and the weights are computed based on a set of theorems provided by Press *et al.*, 1986; page 124).

The estimation was carried out using the GAUSS programming language on a personal computer. We used a high order of integration in the quadrature formulas to evaluate the integrals in the log-likelihood function accurately (we document the high level of accuracy of the quadrature method later in the paper). Gradients of the “quadrature-evaluated” log-likelihood function with respect to the parameters were coded.

6. Data and Empirical Results

The data used in the present study draws from a 1989 Rail Passenger Review conducted by VIA Rail (the Canadian national rail carrier) to develop travel demand models to forecast future intercity travel and estimate shifts in mode split in response to a variety of potential rail service improvements (including high-speed rail) in the Toronto-Montreal corridor (see KPMG Peat Marwick & Koppelman, 1990 for a detailed description of this data). Travel surveys were conducted in the corridor to collect data on intercity travel by four modes (car, air, train and bus). This data included socio-demographic and general trip-making characteristics of the traveler, and detailed information on the current trip (purpose, party size, origin and destination cities, *etc.*). The set of modes available to travelers for their intercity travel was determined based on the geographic location of the trip. Level of service data were generated for each available mode and each trip based on the origin/destination information of the trip.

In this paper, we focus on intercity mode choice for paid business travel in the corridor. The study is confined to a mode choice examination among air, train, and car due to the very few number of individuals choosing the bus mode in the sample and also because of the poor quality of the bus data (see Forinash, 1992). This is not likely to affect the applicability of the mode choice model in any serious way since the bus share for paid business travel in the corridor is less than one percent. The sample used in this study comprises 2769 business travelers. The sample has been weighted to reflect market travel volumes and mode shares.

We estimated five different models in the study: a multinomial logit model, three possible nested logit models, and the heteroscedastic extreme value model. The three nested logit models were: (a) car and train (slow modes) grouped together in a nest which competes against air; (b) train and air (common carriers) grouped together in a nest which competes against car; and (c) air and car grouped together in a nest which competes against train. Of these three structures, the first two seem intuitively plausible, while the third does not. However, we estimate the third structure too because previous research suggests that non-intuitive structures may provide better empirical results (Daly, 1987).

A number of different variable specifications were examined to determine the preferred utility function specification. We arrived at the final specification based on a systematic process of eliminating variables found to be insignificant in previous specifications and based on considerations of parsimony in representation. Among the specifications examined and rejected because they did not provide significantly better results were: (a) out-of-vehicle travel time segmentation into access time (to airport or railway station) and terminal time (at airport or railway station)[3]; (b) travel cost deflated by income to reflect a smaller marginal cost effect on high income travelers than low income travelers (see Ben-Akiva and Lerman, 1985); (c) differential sensitivities of high income and low income groups to changes in in-vehicle and out-of-vehicle travel times and travel cost (see Forinash, 1992); and (d) alternative transformations of the frequency variable (see Peat Marwick Main and Company & Koppelman, 1989).

The final estimation results are shown in Table 1 for the multinomial logit model, the "Heteroscedastic" model imposing the constraints that all the scale parameters are equal to one (we estimate such a model to assess the accuracy of the quadrature procedure used to evaluate the integrals; the results from this model and the multinomial logit model should be the same if the integrals can be evaluated exactly), the nested logit model with car and train grouped as ground modes, and the heteroscedastic model. The estimation results for the other two nested logit models are not shown because the logsum parameter exceeded one in these specifications. This is inconsistent with stochastic utility maximization (McFadden, 1979; Daly & Zachary, 1979).

A comparison of the multinomial logit model and the "Heteroscedastic" model with all scale parameters constrained to one shows that the parameter estimates, their standard errors[4], and the log likelihood function value at convergence are close to each other in the two models. This is an indication of the high level of accuracy of the quadrature method. A comparison of the nested logit model with the multinomial logit model using the likelihood ratio test indicates that the nested logit model fails to reject the multinomial logit model. However, a likelihood ratio test between the heteroscedastic extreme value model and the multinomial logit strongly rejects the multinomial logit in favor of the heteroscedastic specification (the test statistic is 16.56 which is significant at any reasonable level of significance when compared to a chi-squared statistic with two degrees of

[3] Out-of-vehicle travel time is zero for the car mode.
[4] The asymptotic covariance matrix of parameters in all estimations is computed as $H^{-1}\Delta H^{-1}$, where H is the hessian and Δ is the cross-product matrix of the gradients (H and Δ are evaluated at the estimated parameter values). This provides consistent standard errors of the parameters when weights are used.

freedom). Table 1 also evaluates the models in terms of the adjusted likelihood ratio index ($\overline{\rho}^2$).[5] These values again indicate that the heteroscedastic model offers the best fit in the current empirical analysis [note that the nested logit model and the heteroscedastic models can be directly compared to each other using the non-nested adjusted likelihood ratio index test proposed by Ben-Akiva and Lerman (1985); in the current case, the heteroscedastic model specification rejected the nested specification using this non-nested hypothesis test].

In the subsequent discussion on interpretation of model parameters, we will focus only on the multinomial logit and heteroscedastic extreme value models. The signs of all the parameters in the two models are consistent with *a priori* expectations (the car mode is used as the base for the alternative specific constants and alternative specific variables). There is a preference for the common carrier modes (air and train) over the car mode, and for the train mode over the air mode, for trips which originate, end, or originate and end at a large city. The income parameters show that higher income favors air travel relative to other modes, and low income favors train travel. All level-of-service measures yield reasonable parameters. The higher negative coefficient on out-of-vehicle travel time relative to that on in-vehicle travel time reveals that out-of-vehicle travel time is more onerous than in-vehicle travel time.

The parameter estimates from the multinomial logit and the heteroscedastic model are close to each other. However, there are some significant differences (all significance tests are conducted at the 0.1 level). The heteroscedastic model suggests a higher positive probability of choice of the train mode for trips which originate, end, or originate and end at a large city. It also indicates a lower sensitivity of travelers to frequency of service and travel cost; *i.e.*, the heteroscedastic model suggests that travelers place substantially more importance on travel time than on travel cost or frequency of service. Thus, according to the heteroscedastic model, reductions in travel time (even with a concomitant increase in fares) may be a very effective way of increasing the mode share of a travel alternative. The implied cost of in-vehicle travel time is $14.70 per hour in the multinomial

---

[5] The adjusted likelihood ratio index is defined as follows:

$$\overline{p}^2 = 1 - \frac{L(M) - K}{L(C)}$$

where $L(M)$ is the model log-likelihood value, $L(C)$ is the log-likelihood value with only alternative specific constants and an IID error covariance matrix, and K is the number of parameters (besides the alternative specific constants) in the model.

logit and $20.80 per hour in the heteroscedastic model. The corresponding figures for out-of-vehicle travel time are $50.20 and $68.30 per hour, respectively.

The heteroscedastic model indicates that the scale parameter of the random error component associated with the train (air) utility is significantly greater (smaller) than that associated with the car utility (the scale parameter of the random component of car utility is normalized to one; the *t*-statistics for the train and scale parameters are computed with respect to a value of one). Therefore, the heteroscedastic model suggests unequal cross-elasticities among the modes.

Table 2 shows the elasticity matrix with respect to changes in rail level of service characteristics (computed for a representative inter-city business traveler in the corridor) for the multinomial logit and heteroscedastic extreme value models.[6] Two important observations can be made from this table. First, the multinomial logit model predicts higher percentage decreases in air and car choice probabilities and a higher percentage increase in rail choice probability in response to an improvement in train level of service than the heteroscedastic model. Second, the multinomial logit elasticity matrix exhibits the IIA property because the elements in the second and third columns are identical in each row. The heteroscedastic model does not exhibit the IIA property; a 1% change in the level of service of the rail mode results in a larger percentage change in the probability of choosing air than auto. This is a reflection of the lower variance of the random component of the utility of air relative to the random component of the utility of car. We discuss the policy implications of these observations in the next section.

7. Summary and Conclusions

This paper has developed a random utility model with independent, but non-identical error terms distributed with a type I extreme value distribution. The resulting heteroscedastic extreme value model has a number of advantages over other commonly used discrete choice models. The paper proposes and applies an efficient gaussian quadrature method to estimate the heteroscedastic extreme value model.

The heteroscedastic model avoids the IIA restriction of the multinomial logit model by allowing the random components of utilities of the different alternatives to have unequal scale

---

[6] Since the objective of the original study for which the data were collected was to examine the effect of alternative improvements in rail level of service characteristics, we focus on the elasticity matrix corresponding to changes in rail level of service here.

parameters. The IIA assumption holds only if the scale parameters of all the alternatives are equal, in which case the heteroscedastic model collapses to the multinomial logit model.

The empirical analysis of the paper applied the heteroscedastic model, the multinomial logit model, and the nested logit model to the estimation of inter-city travel mode choice in the Toronto-Montreal corridor. The results of the heteroscedastic model reject the multinomial logit formulation of mode choice; corresponding nested logit formulations, however, are either inconsistent with utility maximization principles or are not significantly better than the multinomial logit model. The heteroscedastic model predicts a higher probability of choice of the train mode compared to the multinomial logit model for trips that originate, end, or originate and end at a large city. It also suggests a lower sensitivity of travelers to frequency of service and travel cost relative to the multinomial logit.

Our findings indicate that using a multinomial logit formulation to examine the effects of improving rail level of service in the context of the Canadian data leads to overestimation in the choice probability of the improved rail mode and overestimation in the decrease in the choice probability of non-rail modes. The multinomial logit model also predicts (incorrectly) that the increase in rail choice probability is due to equal percentage decreases in the choice probabilities of the non-rail modes (at the individual level). The results of the heteroscedastic extreme value model show a larger percentage decrease in the air mode than the auto mode.

The observations made above have important policy implications at the aggregate level (these policy implications are specific to the Canadian context; caution must be exercised in generalizing the behavioral implications based on this single application). *First*, the results indicate that the increase in rail mode share in response to improvements in the rail mode is likely to be substantially lower than what might be expected based on the multinomial logit formulation. Thus, the multinomial logit model overestimates the potential ridership on a new (or improved) rail service and, therefore, overestimates revenue projections. This finding is very important, particularly in the light of a recent study which indicates that many public rail investments have suffered huge financial setbacks because they were based on ridership forecasts which later proved to be substantially overestimated (Transportation Systems Center, 1989). *Second*, our results indicate that the potential of an improved rail service to alleviate auto-traffic congestion on intercity highways and air-traffic congestion at airports is likely to be lesser than that suggested by the multinomial logit model. This finding has a direct bearing on the evaluation of alternative strategies to alleviate intercity travel congestion. *Third*, the differential cross-elasticities of air and auto modes in the heteroscedastic logit model suggests that an improvement in the current rail service will alleviate air-traffic congestion at airports more so than alleviating auto-congestion on roadways. Thus, the potential benefit from improving the rail service will depend on the situational context; that is, whether the thrust of the congestion-alleviation effort is to reduce roadway congestion or to reduce air traffic congestion. These findings point to the deficiency of the multinomial logit model as a tool to making informed policy decisions to alleviate intercity travel congestion.

It is important to emphasize that the results obtained in this paper are specific to the current analysis. In general, the heteroscedastic model proposed here is likely to be superior to the multinomial logit and nested logit models in cases with nonidentical-independent random components. However, the nested logit model may be superior to the heteroscedastic and multinomial logit models in cases with identical-nonindependent random components. Finally, note that both the nonidentical-independent and identical-nonindependent cases may be approximations to the nonidentical-nonindependent case and it would be useful to examine the conditions under which the nonidentical-independent case is a better approximation than the identical-nonindependent case and *vice versa*. This is an area for future research.

Acknowledgements

The author would like to thank Chris Forinash for providing the data used in estimation and Prof. Michael Zazanis for helping with the proof in Appendix A. Professor Frank Koppelman and two anonymous referees provided valuable comments and suggestions on an earlier draft of the paper. This research was funded, in part, by a University of Massachusetts Faculty Research Initiation Grant.

References

Allenby, G.M. and Ginter J.L. (1993) Modeling competitive subsets and product differentiation. Research Report, The Ohio State University, Columbus, Ohio.

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

Bunch, D.S. and Kitamura, R. (1990) Multinomial probit model estimation revisited: testing estimable model specifications, maximum likelihood algorithms, and probit integral approximations for trinomial models of household car ownership. Unpublished Working Paper, Graduate School of Business, University of California, Davis.

Butler, J.S. and Moffitt R. (1982) A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica* **50**, 761-764.

Borsh-Supan, A. (1987) *Econometric Analysis of Discrete Choice*. Lecture notes in economics and mathematical systems, Springer, Berlin.

Chintagunta P.K., Jain D.C. and Vilcassim N.J. (1991) Investigating heterogeneity in brand preferences in logit models for panel data. *J. Mktng. Res.* **28**, 417-428.

Currim, I.S. (1982) Predictive testing of consumer choice models not subject to independence of irrelevant alternatives. *J. Mktng. Res.* **19**, 208-222.

Daganzo, C. (1979) *Multinomial Probit: The Theory and its Application to Demand Forecasting*, Academic Press, New York.

Daly, A. (1987) Estimating tree logit models. *Transpn Res.* **21B**, 4, 251-267.

Daly, A. and Zachary, S. (1979) Improved multiple choice models. *Identifying and Measuring the Determinants of Mode Choice*. (Hensher. D. and Davi Q. Eds). Teakfield, London.

Federal Aviation Administration Report (1987) *Terminal Area Forecasts 1987-2000*. U.S. Department of Transportation.

Forinash, C.V. (1992) A Comparison of model structures for intercity travel mode choice. Unpublished Master's Thesis, Department of Civil Engineering, Northwestern University, Evanston, Illinois.

Forinash, C.V. and Koppelman F.S. (1993) Application and interpretation of nested logit models of intercity mode choice. *Transpn Res. Rec.* **1413**, 98-106.

Horowitz, J.L. (1981) Identification and diagnosis of specification errors in the multinomial logit model *Transpn Res.* **15B**, 345-360.

Horowitz, J.L. (1991) Reconsidering the multinomial probit model. *Transpn Res.* **25B**, 433-438.

Johnson, N. and Kotz, S.(1970) *Distributions in Statistics: Continuous Univariate Distributions (Chapter 21).* Wiley, New York.

Koppelman, F.S., Adler T.J., Williams B. and Bhat C.R. (1993) Estimation and testing of alternative approaches to model intercity rail ridership. *1st Int. Conf. on High Speed Ground Transportation Systems*, Florida.

KPMG Peat Marwick and Koppelman F.S. (1990) Proposal for analysis of the market demand for high speed rail in the Quebec/Ontario corridor. Submitted to Ontario/Quebec Rapid Train Task Force.

KPMG Peat Marwick in association with ICF Kaiser Engineers, Inc., Midwest System Sciences, Resource Systems Group, Comsis Corporation and Transportation Consulting Group (1993) Florida high speed and intercity rail market and ridership study: final report. Submitted to Florida Department of Transportation.

McFadden, D. (1973) Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics* (Zaremmbka P. Ed.). Academic Press, New York.

McFadden, D. (1978) Modelling the choice of residential location. *Spatial Interaction Theory and Residential Location* (Karlquist A. Ed., pp. 75-96). North Holland, Amsterdam.

McFadden, D. (1979) Quantitative methods for analyzing travel behavior of individuals: some recent developments. *Behavioral Travel Modelling* (Hensher D. and Stopher P. Eds.). Croom Helm, London.

Moon, F.B. (1991) Houston moves forward again, *Civ. Engng* **61**, 40-43.

Peat Marwick Main and Company and Koppelman F.S. (1989) Preparation of base year service characteristics and development of travel demand models: working paper 3. Report submitted to Ohio High Speed Rail Authority.

Press, W.H., Flannery, B.P., Teukolsky S.A. and Vetterling W.T. (1986) *Numerical Recipes*. Cambridge University Press, Cambridge.

Stopher, P.R., Meyburg A.H. and Brög W. (1981) Travel behavior research: a perspective. *New Horizons in Travel Behavior Research* (Stopher, P.R., Meyburg A.H. and Brög W. Eds.). Lexington Books, Massachusetts.

Transportation Research Board Special Report (1991) In pursuit of speed: new options for intercity passenger transport. 233, National Research Council, Washington, D.C.

Transportation Systems Center (1989) Urban rail transit projects: forecast versus actual ridership and cost. U.S. Department of Transportation, Cambridge.

U.S. Army Corps of Engineers (1990) Preliminary implementation plan, National Maglev Initiative.

Appendix A: Sum of Choice Probabilities in Heteroscedastic Model

In this appendix, we prove that the choice probabilities of alternatives sum to one in the heteroscedastic extreme value model. For ease in presentation, we will establish this result for the case of three alternatives. Generalization to a different number of alternatives is straight-forward.

For the three alternative cases, we can write the choice probabilities based on equation (5) of the text as

$$
\begin{aligned}
P_1 &= \int_{z=-\infty}^{z=+\infty} \Lambda\left[\frac{V_1 - V_2 + \theta_1 z}{\theta_2}\right]\Lambda\left[\frac{V_1 - V_3 + \theta_1 z}{\theta_3}\right]\lambda(z)dz \\
P_2 &= \int_{z=-\infty}^{z=+\infty} \Lambda\left[\frac{V_2 - V_1 + \theta_2 z}{\theta_1}\right]\Lambda\left[\frac{V_2 - V_3 + \theta_2 z}{\theta_3}\right]\lambda(z)dz \qquad \text{(A.1)} \\
P_3 &= \int_{z=-\infty}^{z=+\infty} \Lambda\left[\frac{V_2 - V_1 + \theta_2 z}{\theta_1}\right]\Lambda\left[\frac{V_2 - V_3 + \theta_2 z}{\theta_3}\right]\lambda(z)dz
\end{aligned}
$$

Let us define a function $H(z)$ as follows

$$
H(z) = \Lambda(z)\Lambda\left[\frac{V_1 - V_2 + \theta_1 z}{\theta_2}\right]\Lambda\left[\frac{V_1 - V_3 + \theta_1 z}{\theta_3}\right]. \qquad \text{(A.2)}
$$

Since $H(z)$ is a product of proper cumulative distribution functions, it is also a proper cumulative distribution function. Thus, we can write

$$
\int_{z=-\infty}^{z=+\infty} \frac{d}{dz}\, H(z)dz = H(+\infty) - H(-\infty) = 1\,. \qquad \text{(A.3)}
$$

Carrying out the differentiation of $H(z)$ with respect to z using (A.2), we can also write:

$$
\begin{aligned}
\int_{z=-\infty}^{z=+\infty} \frac{d}{dz}\, H(z)dz = &\int_{z=-\infty}^{z=+\infty} \lambda(z)\,\Lambda\left[\frac{V_1 - V_2 + \theta_1 z}{\theta_2}\right]\Lambda\left[\frac{V_1 - V_3 + \theta_1 z}{\theta_3}\right]dz \quad + \\
&\int_{z=-\infty}^{z=+\infty} \Lambda(z)\,\lambda\left[\frac{V_1 - V_2 + \theta_1 z}{\theta_2}\right]\Lambda\left[\frac{V_1 - V_3 + \theta_1 z}{\theta_3}\right]\left(\frac{\theta_1}{\theta_2}dz\right) \quad + \qquad \text{(A.4)} \\
&\int_{z=-\infty}^{z=+\infty} \Lambda(z)\,\Lambda\left[\frac{V_1 - V_2 + \theta_1 z}{\theta_2}\right]\lambda\left[\frac{V_1 - V_3 + \theta_1 z}{\theta_3}\right]\left(\frac{\theta_1}{\theta_3}dz\right).
\end{aligned}
$$

Making the change of variables; $z'' = (V_1 - V_3 + \theta_1 z)/\theta_3$ in the second integral term and in the $z' = (V_1 - V_2 + \theta_1 z)/\theta_2$ in the third integral term; it is easy to see that the first term in equation (A.4) is P_1, the second is P_2, and the third is P_3. Thus, by (A.3), $P_1 + P_2 + P_3 = 1$.

List of Tables

Table 1. Intercity Mode Choice Estimation Results

Variable	Multinomial Logit		Constrained "Hetero-scedastic" Model		Nested Logit with Car and Train Grouped		Heteroscedastic Extreme Value Model	
	Parameter	t-statistic	Parameter	t-statistic	Parameter	t-statistic	Parameter	t-statistic
Mode Constants (car is base)								
Train	-0.5396	-1.55	-0.5572	-1.66	-0.6703	-2.14	-0.1763	-0.42
Air	-0.6495	-1.23	-0.6055	-1.09	-0.5135	-1.31	-0.4883	-0.88
Large City Indicator (car is base)								
Train	1.4825	7.98	1.3971	8.28	1.3250	6.13	1.9066	6.45
Air	0.9349	5.33	0.9370	5.49	0.8874	5.00	0.7877	4.96
Household Income (car is base)								
Train	-0.0108	-3.33	-0.0106	-3.34	-0.0101	-3.30	-0.0167	-3.57
Air	0.0261	7.02	0.0258	7.02	0.0262	7.42	0.0223	6.02
Frequency of service	0.0846	17.18	0.0835	17.46	0.0846	17.67	0.0741	10.56
Travel Cost	-0.0429	-10.51	-0.0420	-10.63	-0.0414	-11.03	-0.0318	-5.93
Travel Time								
In-Vehicle	-0.0105	-13.57	-0.0102	-13.64	-0.0102	-12.64	-0.0110	-9.78
Out-of-Vehicle	-0.0359	-12.18	-0.0349	-12.43	-0.0353	-13.86	-0.0362	-8.64
Logsum Parameter[1]	1.0000	-	1.0000	-	0.9032	1.14	1.0000	-
Scale Parameters (car parm.= 1)[2]								
Train	1.0000	-	1.0000	-	1.0000	-	1.3689	2.60
Air	1.0000	-	1.0000	-	1.0000	-	0.6958	2.41
Log Likelihood At Convergence[3]	-1828.89		-1830.10		-1828.35		-1820.60	
Adjusted Likelihood Ratio Index	0.3525		0.3521		0.3524		0.3548	

[1] The logsum parameter is implicitly constrained to one in the multinomial logit and heteroscedastic model specifications. The t-statistic for the logsum parameter in the nested logit is with respect to a value of one.

[2] The scale parameters are implicitly constrained to one in the multinomial logit and nested logit models and explicitly constrained to one in the constrained "heteroscedastic" model. The t-statistics for the scale parameters in the heteroscedastic model are with respect to a value of one.

[3] The log likelihood value at zero is -3042.06 and the log likelihood value with only alternative specific constants and an IID error covariance matrix is -2837.12.

Table 2. Elasticity Matrix in Response to Change in Rail Service for Multinomial Logit and Heteroscedastic Models

Rail Level of Service Attribute	Multinomial Logit Model			Heteroscedastic Extreme Value Model		
	Train	Air	Car	Train	Air	Car
Frequency	0.303	-0.068	-0.068	0.205	-0.053	-0.040
Cost	-1.951	0.436	0.436	-1.121	0.290	0.220
In-Vehicle Travel Time	-1.915	0.428	0.428	-1.562	0.404	0.307
Out-of-Vehicle Travel Time	-2.501	0.559	0.559	-1.952	0.504	0.384

Note: The elasticities are computed for a representative intercity business traveler in the corridor.