Application and Interpretation of Nested Logit Models of Intercity Mode Choice

CHRISTOPHER V. FORINASH AND FRANK S. KOPPELMAN

A clear understanding of the sources and amount of ridership on a new or improved travel mode is critical to evaluating the financial, travel flow, and external impacts of proposed improvements. The multinomial logit model traditionally used to model intercity mode choice may not adequately reflect traveler behavior because it restricts the relative probability of choosing between any pair of existing modes to be unchanged when other modes are introduced or changed. The nested logit model provides a computationally feasible generalization to the multinomial logit model, which allows for specified mode pairs to exhibit increased sensitivity to changes in service. Full information estimation of nested logit models allows efficient use of information and yields results directly comparable to multinomial logit models. Business travel in the Ontario-Quebec corridor of Canada is examined. A set of nested logit structures that allow for various combinations of differential sensitivity to changes in service quality of rail is estimated. Nested logit structures with bus-train or car-train nests prove superior to the multinomial logit model. Both of the nested logit models predict larger increases in rail shares than the multinomial logit model in response to rail service improvements, but the source of that increased ridership differs between the nested logit structures. This points to the need for models of individual choice that retain the advantages of nested logit while allowing pairwise similarity between alternatives.

Congestion in intercity travel increases the cost of travel directly through the loss of traveler time and indirectly through increased costs in system operation. These costs are transferred to travelers and others by common carriers through fares and by governments through taxes or debt. Considerable attention has been directed toward rapidly increasing congestion during the last decade and projections of substantial additional increases through the next two decades (*1*,*2*). Proposals to alleviate existing and projected congestion include construction of new airports (*3–5*); construction or widening of express highways, some with toll charges (*6–8*); upgrading of conventional rail services (*9*,*10*) and construction of new high-speed ground transportation based on rail or magnetic levitation technology (*11*). It has been difficult to implement many of these proposals because of concerns about financing and environmental impacts and differences among governmental and private institutions. The difficulty reaching positive implementation decisions for both new airport and high-speed ground transportation alternatives may, in part, be due to concerns about the quality of ridership and revenue forecasts.

A fundamental issue in the prediction of ridership is the ability to model and explain the likely projected changes in ridership and the sources of projected ridership. A clear representation of the sources of new ridership on new or improved alternatives can increase the confidence of both public and private investors in the likelihood of recovering their investment. It can also be used to scale the beneficial effect of the investment on congestion and external impacts. This paper tests the application of the nested logit model to estimate ridership on intercity travel modes and compares the results of the nested logit model to the more commonly used multinomial logit model. The issue of predicting changes in total ridership in response to improvements in modal service is not addressed in this paper but has been addressed by others (*12*,*13*).

The multinomial logit model has been used almost exclusively to model both urban and intercity mode choice until recently (*14*,*15*). The multinomial logit model is widely used because its mathematical form is simpler than that of alternative models, making it easier to estimate and interpret. However, the important disadvantage of the multinomial logit model is that it restricts the relative probability of choosing between any pair of unchanged modes to be unchanged due to changes in other modes of travel. This restriction implies that the introduction of any new mode or the improvement of any existing mode will affect all other modes proportionally. This property of equal proportional change or equal cross-elasticity of unchanged modes is unlikely to represent actual choice behavior in a variety of situations. Such misrepresentation of choice behavior can result in incorrect estimated models and incorrect predictions of mode share and diversion from existing modes. Differences in the impact of the introduction of new services on existing modes can be addressed by adoption of the multinomial probit model, which is rarely used in application due to problems of complexity, estimation, and interpretation, or the nested logit model.

Studies of intercity mode choice that have used the multinomial logit model include the Ontario-Quebec corridor in Canada (*12*), Twin Cities-Duluth in Minnesota (*16*), and the United States as a whole (*17–19*). Although the nested logit model was recommended for "immediate implementation" at the 3rd International Conference on Behavioural Travel Modeling in 1977 (*14*), it use has been limited due, in part, to the limited availability of the more flexible software needed to estimate the nested logit model relative to the availability of a variety of software to estimate the multinomial logit model. The nested logit model has been used to estimate mode choice models for urban mode-choice and for multimodal and multidimensional choices (*20–23*), although the older efforts were accomplished using inefficient two-stage limited-information maximum likelihood estimation. Hensher (*15*) recommended adoption of the nested logit model for inter-

The Transportation Center, 1936 Sheridan Road, Evanston, Ill. 60208.

city mode choice estimation. However, there have been few applications of the nested logit model in the intercity mode choice context. These include the estimation of a multidimensional mode, destination, and rental-car choice model (*24*) and a nested mode and air-fare-class choice model (Koppelman, unpublished data), both using limited-information estimation.

NESTED LOGIT MODEL DESCRIPTION AND PROPERTIES

The nested logit and multinomial logit models can each be depicted by a tree structure that represents all the alternatives. The multinomial logit model treats all alternatives equally, whereas the nested logit model includes intermediate branches that group alternatives (Figure 1). The grouping of alternatives indicates the degree of sensitivity (cross-elasticity) among alternatives. Alternatives in a common nest show the same degree of increased sensitivity compared to alternatives not in the nest. Thus, although the nested logit model is not completely flexible in the sense that distinct pairwise sensitivities cannot be estimated, it provides a more general structure than the multinomial logit model. The differences in structure can result in dramatically different mode ridership projections and diversions than those obtained by the multinomial logit model in cases where the nested logit model is significantly different from the multinomial logit model.

The widely adopted paradigm of utility maximization provides a link by which choice probabilities can be estimated given characteristics of the modes and the decision maker. This paradigm holds that an individual acts to maximize his or her utility by choosing among the available alternatives. Utility can then be estimated as a function of the traveler and mode characteristics. The choice probabilities can be computed as functions of the relative utilities among alternatives. Conventionally, the utility of an alternative, U_{ij}, is assumed to be the sum of a deterministic component, V_{ij}, which describes the characteristics of individual i and the attributes of alternative j, and a random term, ϵ_{ij}, which represents elements not measured or included in the model:

$$U_{ij} = V_{ij} + \epsilon_{ij} \tag{1}$$

Further, the measured and included component of the model is represented by a linear additive function that includes parameters, β, and variables, X_{ij}, which are predetermined functions of the characteristics of individual i and the attributes of alternative j:

$$U_{ij} = \beta' X_{ij} + \epsilon_{ij} \tag{2}$$

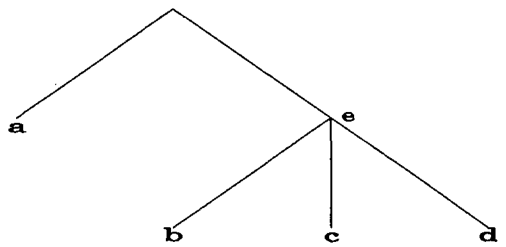

FIGURE 1 Example four-mode nested choice structure: modes *a*, *b*, *c*, and *d*.

Assumptions about the distribution of the error terms ϵ_{ij} lead to different model structures.

The assumption that the error terms are distributed independently and identically over individuals and alternatives, with a Gumbel (0,1) distribution, yields the multinomial logit model (*25,26*):

$$P_{ij} = \frac{\text{esp}(V_{ij})}{\sum_{j' \epsilon J} \exp(V'_{ij})} \tag{3}$$

where J is the set of available alternatives.

The nested logit model is derived from an assumption that some of the alternatives share common components in the random term. That is, the random term, ϵ_j, ignoring the individual subscript for simplicity of notation, can be decomposed into a portion associated with each alternative and a portion associated with groups of alternatives. For example, consider the nested model in Figure 1, where alternatives b, c, and d are included in the nest, which is labeled e. The total errors for alternatives b, c, and d are defined as $\epsilon_b + \epsilon_e$, $\epsilon_c + \epsilon_e$ and $\epsilon_d + \epsilon_e$. The total error for alternative a not in the nest is ϵ_a. The included and measured portion of utility may also be decomposed into two parts representing specific characteristics of the alternative, V_b, V_c, and V_d, and common characteristics of the nested alternatives, V_e. That is:

$$\begin{aligned} U_a &= V_a + \epsilon_a \\ U_b &= V_b + V_e + \epsilon_b + \epsilon_e \\ U_c &= V_c + V_e + \epsilon_c + \epsilon_e \\ U_d &= V_d + V_e + \epsilon_d + \epsilon_e \end{aligned} \tag{4}$$

The nested logit model is obtained by assuming further that the error terms for each alternative—ϵ_a, $\epsilon_b + \epsilon_e$, $\epsilon_c + \epsilon_e$ and $\epsilon_d + \epsilon_e$—are *distributed Gumbel* (0,1) and that the independent portion of the error terms for the nested alternatives—ϵ_b, ϵ_c and ϵ_d—are *distributed independent Gumbel* $(0,\theta^1)$ (*25*). The common error component, ϵ_e, for the nested alternatives represents a covariance relationship that describes an increased similarity between pairs of nested alternatives and leads to a higher sensitivity (cross-elasticity) between alternatives. If this common component, ϵ_e, is reduced to zero, the model reduces to the multinomial logit model with no covariance of error terms among the alternatives.

These assumptions yield the following conditional choice probability for each nested alternative n among the nested alternatives (conditional on choice of the nest at the higher level):

$$P_{n/e} = \frac{\exp(V_n/\theta)}{\exp(V_b/\theta) + \exp(V_c/\theta) + \exp(V_d/\theta)} \tag{5.1}$$

The marginal choice probabilities for alternative a and for the nest are:

$$P_a = \frac{\exp(V_a)}{\exp(V_a) + \exp(V_e + \theta\Gamma_e)} \tag{5.2}$$

$$P_e = \frac{\exp(V_e + \theta\Gamma_e)}{\exp(V_a) + \exp(V_e + \theta\Gamma_e)} \tag{5.3}$$

where Γ_e measures the expected maximum utility among the nested alternatives and is given by the log sum of the exponents of the nested utilities:

$$\Gamma_e = \ln[\exp(V_b/\theta) + \exp(V_c/\theta) + \exp(V_d/\theta)] \tag{5.4}$$

The parameter of the log sum variable, θ, is the estimator of the scale parameter of the Gumbel distribution for the nested alternatives. The probability of choosing any lower-nest alternative n is the product of the probability of the nest being chosen and the conditional probability of that alternative, $P_e \times P_{n/e}$.

The sensitivity of each alternative to changes in other alternatives can be represented by the cross-elasticity, the proportional effect on the probability of choosing alternative j' of a change of an attribute of alternative j. For the multinomial logit model, the cross-elasticities for all pairs of alternatives j and j' are given in Table 1. The equal proportional effect of the introduction of a new alternative or a change in an existing alternative j on all other alternatives is indicated by the lack of dependence of the elasticity on the affected alternative, j'. The self-elasticity for any alternative j is also given in Table 1.

The corresponding elasticities for the nested logit model are differentiated between alternatives that are or are not in the same nest. Using the example of Figure 1 and Equation 5, the effect of a change in one of the nested alternatives, for example, b, on the nonnested alternative a, given in the first line of Section [iii] in Table 1, is identical to that for the multinomial logit case. An identical relationship holds for changes in nonnested alternatives, as shown in Section [ii] in Table 1. However, the corresponding equation for another nested alternative, for example, c, is quite different, as shown in the last line of Section [iii] in Table 1. If θ equals one, its maximum value, the cross-elasticity collapses to the corresponding equation for the multinomial logit model and for the alternatives not in the nest. If θ is between zero and one, as expected, the magnitude of the cross-elasticity for the nested alternatives will be greater than that for the alternatives not in the nest, and greater than that which would be obtained for the multinomial logit model, if the level-of-service parameters do not change. The estimation results in this study produced only small changes in the level-of-service parameters.

The direct-elasticity for any nonnested alternative is identical to that for the multinomial logit model. However, for nested alternatives, the direct-elasticity is as shown in the middle line of Section [iii] in Table 1. Thus, if θ equals one, this equation reduces to that for the multinomial logit model and is the same as for nonnested alternatives. However, for θ less than one, the direct-elasticity is greater than that for the nonnested alternatives (and for the multinomial logit model if the level-of-service parameters are unchanged).

ESTIMATION OF THE NESTED LOGIT MODEL

Estimation of the nested logit model has been most generally undertaken by limited information, maximum likelihood techniques (*21,27*). This method first estimates parameters for the lowest nest(s) and then estimates parameters for successively higher nests based on the computation of the log sum values, which are obtained from the lower nest estimation results (*25*).

This sequential estimation leads to a suboptimal log-likelihood at convergence and can yield a lower log-likelihood than the multinomial logit model (*27–29*). Although the parameter estimates are consistent, they are not efficient and have been found to be quite far from full-information estimates in practice (*15,27,29*).

Estimation of nested logit structures by full-information, maximum likelihood allows the most efficient use of available information. Full-information, maximum likelihood will indicate clearly whether the multinomial logit model can be rejected by the data. Further, constraints can be imposed

TABLE 1 Analytic Elasticities from Multinomial Logit and Nested Logit Models

Elasticity of Probability of Choosing Mode	Mode for Which Level-of-Service Changes: Multinomial Logit Model, Mode j	Nested Logit Model: Mode a not in nest e	Nested Logit Model: Mode b in nest e
Multinomial Logit Model:	[i]		
Mode j	$[1-P_j]\,\beta_{LOS}LOS_j$		
Mode j′	$-P_j\beta_{LOS}LOS_j$		
Nested Logit Model:		[ii]	[iii]
Mode a not in Nest e		$[1-P_a]\,\beta_{LOS}LOS_a$	$-P_b\beta_{LOS}LOS_b$
Mode b in Nest e		$-P_a\beta_{LOS}LOS_a$	$[(1-P_e)\,P_{b/e}+\frac{1}{\theta}(1-P_{b/e})]\,\beta_{LOS}LOS_b$
Mode c in Nest e		$-P_a\beta_{LOS}LOS_a$	$-[P_b+\frac{(1-\theta)}{\theta}P_{b/e}]\,\beta_{LOS}LOS_b$

across nests, unlike in limited-information, maximum likelihood estimation. Parameters and standard errors obtained by full-information, maximum likelihood estimation are also directly comparable to multinomial logit results, unlike those produced by limited-information, maximum likelihood. As computing speed and software have advanced, full-information, maximum likelihood has become feasible and should replace limited-information, maximum likelihood in practice.

Maximum likelihood techniques estimate parameter values by maximizing the likelihood function of a sample. The log of this likelihood function is of the form

$$\mathcal{L} = \sum_{i} wt_i \sum_{j} \delta_{ij} \ln P_{ij} \tag{6}$$

where δ_{ij} equals one if individual i chooses alternative j and zero otherwise, and P_{ij} is the model-based probability that individual i chooses alternative j. Wt_i represents the sample weight on each observation; the sample weights are normalized to sum to the sample size.

The likelihood function for the example nested logit model of Figure 1 and Equation 5 is:

$$\mathcal{L} = \sum_{i} wt_i \sum_{j} \delta_{ij} \ln P_{ij}$$

$$\mathcal{L} = \sum_{i} wt_i \sum_{j} \delta_{ij} (\ln P_{ij} + \ln P_{i,j/e})$$

$$\mathcal{L} = \sum_{i} wt_i \left(\sum_{j=a,e} \delta_j \ln P_j + \sum_{k=b,c,d} gd_k \ln P_{k/e} \right) \tag{7}$$

where δ_j equals one for mode a, if chosen, or δ_j equals one for composite alternative e if any of the modes b, c, or d are chosen; δ_k equals one for the nested alternative, if any, which is chosen. Generally, the likelihood function is the sum of the likelihoods, jointly estimated, for all of the nests in the model.

In full-information estimation, all data are used to estimate all parameters jointly in a single maximum-likelihood procedure. The hessian of the log likelihood function for a nested logit model is not globally concave, unlike that for multinomial logit, and thus convergence to a global maximum is not guaranteed. Thus, optimization of the nested-logit log-likelihood function may need to be performed several times with distinct starting values to increase the chance of locating a global optimum.

Several drawbacks of limited-information, maximum likelihood estimation of nested logit structures demonstrate the preferability of full-information techniques. Because only observations choosing one of the lower-level alternatives can be used in lower-nest estimation in limited-information estimation, the procedure makes inefficient use of the data. In addition, individuals having only one of the lower-nest alternatives available are not used in the first step of estimation, as they do not face a choice at this level.

Another weakness of this procedure is that generic parameters applicable to variables in lower and upper nests must be constrained in the upper nests to the values found in the lower nest, adjusted by the inclusive-value parameter θ. Because the lower nest is estimated with only a subset of the data, this can propagate seriously inefficient estimates through the model structure. Alternative- or nest-specific parameters can be estimated in lieu of imposing equality constraints for level-of-service parameters among nests, but this yields results not directly comparable to multinomial logit with generic parameters.

For upper-nest estimation, Γ is computed on the basis of the parameters estimated in the first step, but the inclusive value Γ is an estimate that includes measurement error. This measurement error is ignored in the higher nest estimation, leading to underestimated uppernest standard errors. This may result in retaining parameters in the model that do not warrant inclusion on statistical grounds. Correction techniques, though included in some new statistical packages, are laborious (*23*).

All results in this paper were obtained with full-information, maximum likelihood estimation, performed by software written by the authors and Dr. Chandra Bhat for this purpose. Because the nested logit likelihood function is not necessarily globally concave, unlike the multinomial logit likelihood function, convergence to a global optima from any starting point is not guaranteed. Estimation starting from the multinomial logit parameter values was found to offer the best chance of convergence to an acceptable value of log likelihood.

ESTIMATION OF MULTINOMIAL AND NESTED LOGIT INTERCITY MODE CHOICE MODEL

The authors applied the nested logit model to the estimation of intercity mode choice for travel in the Ontario-Quebec corridor from Windsor in the west to Quebec City in the east. The data used in this study were assembled by VIA Rail (the Canadian national rail carrier) in 1989 to estimate the demand for high-speed rail in the Toronto-Montreal corridor and support future decisions on rail service improvements in the corridor (*12*). This corridor encompasses several thousand square kilometers of two provinces containing the highest population densities in Canada. The main source of data for the four intercity travel modes of interest (train, air, bus, and car) was a 1989 Rail Passenger Review conducted by VIA Rail. These data include travel volumes and impedance data by mode and travel surveys collected on all four modes in 1988 for travel beginning and ending in 136 districts in the region. For this study, only paid business travel is considered. The 4,324 individual trips in this data set have been weighted by demographic and travel characteristics to reflect more than 20 million annual business trips in the corridor (*12*; Forinash, unpublished data).

The final utility function specification employed in the Ontario-Quebec study is adopted as the base model specification, and improvements to it are considered. The Ontario-Quebec specification includes mode-specific constants and large city variables, and generic frequency, travel cost, and in-vehicle and out-of-vehicle travel times (Model 1 in Table 2) (*12*). Both the in-vehicle and out-of-vehicle travel time components are segmented by annual household income, with the break point at C$30,000 to reflect differences in value-of-time between low- and high-income travelers. This specification obtained significant estimates of all parameters, except the bus-specific large city indicator, and a likelihood ratio of 0.80. The implied values of in-vehicle time are C$25 for high-income travelers and C$7 for low-income travelers; the values of out-

TABLE 2 Utility Function Specification Improvements

Variable Description	Estimated Parameter, T-statistic vs. Zero							
	1. Base Specification		2. With Alternative Income Variables		3. With Modified Time Variables		4. With Income and Modified Time Variables	
Mode Constants								
AIR	1.888	2.8	0.07286	0.1	1.308	1.9	-0.5652	-0.7
BUS	-2.756	-5.7	-1.181	-1.9	-2.705	-5.6	-1.135	-1.8
CAR	2.203	5.8	1.621	3.2	1.312	3.2	0.6588	1.2
TRAIN (Base)	0		0		0		0	
Large City Indicator								
AIR	-0.7460	-3.6	-0.7843	-3.8	-0.7244	-3.5	-0.7611	-3.6
BUS	-0.1224	-0.4	0.07158	0.2	-0.06979	-0.2	0.1214	0.4
CAR	-1.306	-7.4	-1.309	-7.5	-1.213	-6.8	-1.214	-6.8
Household Income								
AIR			0.03157	3.7			0.03290	3.9
BUS			-0.04308	-3.8			-0.04304	-3.8
CAR			0.01007	1.4			0.01093	1.5
Frequency	0.1022	13.5	0.1007	13.3	0.1018	13.8	0.1005	13.7
Travel Cost	-0.03265	-5.2	-0.03128	-4.9	-0.02514	-4.1	-0.02391	-3.8
Travel Time								
In-Vehicle								
High Income	-0.01382	-10.3	-0.01315	-9.7				
Low Income	-0.003797	-2.1	-0.00952	-4.4				
Out-of-Vehicle								
High Income	-0.04053	-9.9	-0.03971	-9.6				
Low Income	-0.02636	-6.0	-0.03399	-6.6				
OVT/log(D)								
High Income					-0.2011	-8.3	-0.2020	-8.2
Low Income					-0.1760	-7.0	-0.1889	-6.8
Total								
High Income					-0.01283	-9.5	-0.01214	-8.9
Low Income					-0.00307	-1.7	-0.00888	-4.2
Log Likelihood								
At Convergence	-1072.1		-1051.9		-1058.4		-1037.7	
At Market Shares	-1951.1		-1951.1		-1951.1		-1951.1	
At Zero	-5334.2		-5334.2		-5334.2		-5334.2	
L'hood Ratio Index								
vs. Market Shares	0.448		0.461		0.458		0.468	
vs. Zero	0.799		0.803		0.802		0.805	

Note: OVT/log(D) = out-of-vehicle travel time over log of the distance in kilometers.

of-vehicle time are C$74 and C$48, for high- and low-income travelers, respectively.

The authors have considered two specification improvements to the Ontario-Quebec model. First, the model could include alternative-specific income variables to reflect the change in average biases for or against each mode due to changes in income. The addition of these variables (Model 2) is highly significant. The travel time variables could also be reformulated to total travel time and out-of-vehicle travel time divided by the log of distance traveled, still segmented by income (Model 3). This modification is also highly significant. Finally, the authors have considered both changes in specification (Model 4) which were adopted as the preferred multinomial logit model.

The preferred model (Model 4) provides a significant improvement in fit relative to each of the other models. Also, each service parameter is significant at the 1 percent level. Of the mode-specific parameters, only the income parameter for car, the large city indicator for bus, and the constants are insignificant at the 1 percent level. These merely indicate, respectively, that the effect on car utility of income is approximately the same as income's effect on train utility, that the utilities of bus and rail increase equally if traveling to or from a large city, and that all modes have approximately equal utility, *ceteris paribus*.

The large-city parameters indicate that each of the common-carrier modes (train, air, and bus) benefit relative to the automobile from having either or both ends of a trip in a population center, with train and bus benefiting more. The income parameters show that higher income favors air travel relative to other modes, and low income favors bus travel. All level-of-service measures available in the data are included and yield reasonable parameters.

The transformation of travel time in the preferred specification constrains the monetary value of out-of-vehicle travel time to equal or exceed that of in-vehicle travel time, with the difference diminishing with increasing trip distance. For shorter trips, travelers are likely to be much more sensitive to differences in access time than run time, but this difference is likely to decrease with trip distance. Similar transformations, based on distance instead of log of distance, have been used in urban mode choice (*25,30,31*). The values of out-of-vehicle and in-vehicle travel time can be derived as

$$VOT_{OVT} = \frac{\beta_{TT} + \dfrac{\beta_{OVT/\log(D)}}{\log(D)}}{\beta_{TC}}$$

$$VOT_{IVT} = \frac{\beta_{TT}}{\beta_{TC}} \qquad (8)$$

where β_{TT} is the parameter for total travel time, $\beta_{OVT/\log D}$ is the parameter for out-of-vehicle time divided by the log of the travel distance, and β_{TC} is the parameter for travel cost. The specification yields similar values of in-vehicle travel time to the Ontario-Quebec specification: C$22 for high-income travelers and C$16 for low-income travelers. Higher values of out-of-vehicle travel time are implied by this model, C$92 and C$83 for high- and low-income travelers, respectively, evaluated at 231 km, the average distance traveled.

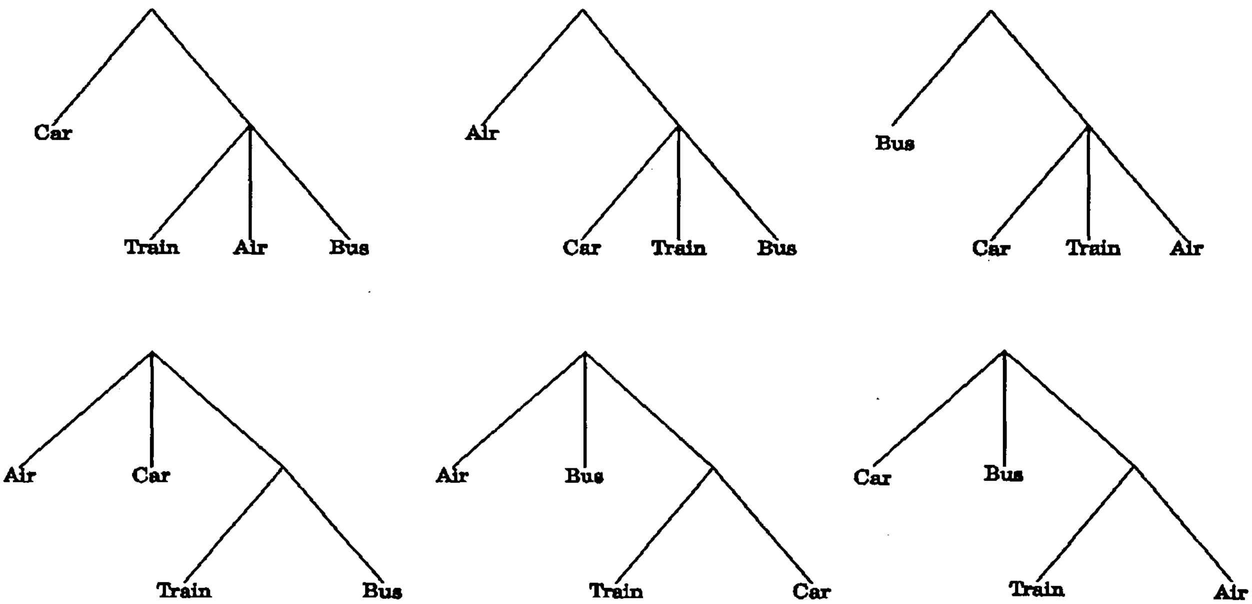

FIGURE 2 Two-level nested choice structures with train nested: modes train, air, bus, and car.

The authors used this specification to estimate alternative nested logit structures. There are 16 two-level and 12 three-level nested logit structures among the four available alternatives. Daly (*21*) found that initial screening on the basis of intuition may eliminate structures that turn out to be statistically superior. This paper considers the six two-level structures that include the rail alternative in the lower nest (Figure 2). These six structures represent various combinations of differential sensitivity to changes in service quality of rail, the mode being considered for service improvement.

Three of these six structures obtained estimates of the log sum parameter that were in the acceptable range and significantly different from one, thus rejecting the multinomial logit model (Table 3). The train-bus nested structure implies higher sensitivity between train and bus than other mode pairs, whereas the train-car nested structure implies higher sensitivity between train and car than other mode pairs. The train-bus-car nested structure includes increased sensitivity between both train and bus and train and car, but also implies increased sensitivity between bus and car, which is not supported by

TABLE 3 Plausible Nesting Structures Revealed by the Data

	Estimated Parameter, T-statistic vs. Zero (vs. Unity for Inclusive Value Parameter)							
Variable Description	Multinomial Logit		a. Train, Bus, Car Nested		b. Train, Bus Nested		c. Train, Car Nested	
Mode Constants								
AIR	-0.5652	-0.7	-1.532	-1.8	-0.7359	-0.9	-1.916	-2.5
BUS	-1.135	-1.8	-1.344	-2.6	-1.3140	-2.6	-2.278	-3.7
CAR	0.6588	1.2	-0.06759	-0.1	0.5756	1.1	-0.3197	-0.7
TRAIN (Base)	0		0		0		0	
Large City Indicator								
AIR	-0.7611	-3.6	-0.4960	-2.3	-0.7697	-3.8	-0.3447	-1.7
BUS	0.1214	0.4	0.1561	0.6	0.1339	0.5	0.5476	1.6
CAR	-1.214	-6.8	-0.9190	-4.9	-1.242	-7.3	-0.7626	-4.7
Household Income								
AIR	0.03290	3.9	0.03022	3.9	0.03507	4.2	0.02899	4.0
BUS	-0.04304	-3.8	-003323	-3.3	-0.03009	-3.2	-0.04750	-4.4
CAR	0.01093	1.5	0.00896	15	0.01303	1.9	0.00749	1.4
Frequency	0.1005	13.7	0.1002	14.2	0.09843	13.4	0.1017	14.6
Travel Cost	-0.02391	-3.8	-0.02079	-3.7	-0.02339	-3.8	-0.02113	-4.1
Travel Time								
OVT/log(D)								
High Income	-0.2020	-8.2	-0.1958	-8.8	-0.1977	-8.1	-0.1866	-8.7
Low Income	-0.1889	-6.8	-0.1872	-7.7	-0.1830	-6.6	-0.1775	-7.6
Total								
High Income	-0.01214	-8.9	-0.01174	-9.1	-0.01217	-9.0	-0.01134	-9.0
Low Income	-0.00888	-4.2	-0.0829	-4.2	-0.00883	-4.2	-0.00815	-4.2
Inclusive Value	1.0		0.788	2.3*	0.6488	2.7*	0.6714	4.0*
Log Likelihood								
At Convergence	-1037.7		-1035.8		-1035.8		-1033.4	
At Market Shares	-1951.1		-1951.1		-1951.1		-1951.1	
At Zero	-5334.2		-5334.2		-5334.2		-5334.2	
L'hood Ratio Index								
vs. Market Shares	0.468		0.469		0.469		0.470	
vs. Zero	0.805		0.806		0.806		0.806	

Notes: OVT/log(D) is out-of-vehicle travel time deflated by the common log of the distance in kilometers.

estimation of a model with bus and car only in the lower nest. Thus, the authors prefer the train-bus and train-car nested structures to the train-bus-car nested structure. The train-car nested structure provides the best fit to the data.

The estimates for the level-of-service parameter estimates for all three structures have the correct sign and are highly significant. Further, these estimates are close to those obtained for the multinomial logit model. Thus, the values of time implied by these models are similar to those reported for the multinomial logit model. The parameter estimates for alternative-specific income variables differ somewhat more but are within one standard error in most cases. The parameter estimates for the alternative-specific constants and large city variables differ considerably among models reflecting the need to adjust these variables to compensate for the changes in model structure.

IMPLICATIONS OF NESTED LOGIT ESTIMATION FOR PREDICTION OF RAIL SHARES

The demonstration that the nested logit model statistically rejects the multinomial logit model provides important and useful insight into the likely behavioral response of travelers to changes in rail travel service. The authors are also interested in the impact of these changes in model structure on the changes in predicted ridership if specific changes in rail service are undertaken in the future. The authors explored this by estimating the differences in mode choice probabilities predicted for representative individuals traveling between specific city pairs. The ridership predictions were prepared using the incremental logit formulations (*32*) of the multinomial logit model, and the nested logit models with train and bus nested and with train and car nested.

Table 4 presents the market size and current (1987) mode shares for three example markets: Ottawa-Toronto, Toronto-Montreal, and Ottawa-Montreal. Adopting the market shares as representative mode choice probabilities and using average values of all variables, the projected mode probabilities for each city pairs based on the multinomial logit model and the two nested logit models are reported in Table 5, for a 40 percent reduction in train in-vehicle travel time. This approximates the improvement high-speed rail offers, boosting the line-haul average speed from around 100 km/hr (62 mph) to about 160 km/hr (100 mph). All three models predict a substantial increase in train probability; however, the increases for the two nested logit models are substantially higher than for the multinomial logit model (except for the Toronto-Montreal pair for the train-bus nest due to the initial zero mode probability for the bus alternative). The increased rail share results from increased shifting from the other nested alternative, bus or car, to the rail alternative. There is little difference in air shares among the models.

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

This paper demonstrated a statistically significant rejection of the multinomial logit model in favor of three alternative nested

TABLE 4 Description of Overall and Sample Markets

Travel Market	Distance (kilometers)	1987 Train Travel Time (minutes)	1987 Market Size (annual business travelers)	1987 Market Shares (%)			
				Train	Air	Bus	Car
Ottawa-Toronto	420	263	459,000	4.87	72.62	1.95	20.56
Toronto-Montreal	540	295	531,000	8.74	81.00	0.00	10.27
Ottawa-Montreal	206	121	601,000	9.09	9.27	4.83	76.81

TABLE 5 Projected Market Shares

<table>
<tr><th rowspan="3">Travel Market</th><th colspan="12">Future Market Shares (%) Predicted with 40% Improvement in Train In-Vehicle Travel Time</th></tr>
<tr><th colspan="3">Train</th><th colspan="3">Bus</th><th colspan="3">Car</th><th colspan="3">Air</th></tr>
<tr><th>Multinomial Logit (MNL)</th><th>Train/Bus Nested Logit (T/B NL)</th><th>Train/Car Nested Logit (T/C NL)</th><th>MNL</th><th>T/B NL</th><th>T/C NL</th><th>MNL</th><th>T/B NL</th><th>T/C NL</th><th>MNL</th><th>T/B NL</th><th>T/C NL</th></tr>
<tr><td>Ottawa-Toronto</td><td>14.89 (+206%)</td><td>16.32 (+235%)</td><td>19.32 (+297%)</td><td>1.74 (-11%)</td><td>0.98 (-50%)</td><td>1.72 (-12%)</td><td>18.39 (-11%)</td><td>18.25 (-11%)</td><td>14.80 (-28%)</td><td>64.97 (-11%)</td><td>64.45 (-11%)</td><td>64.16 (-12%)</td></tr>
<tr><td>Toronto-Montreal</td><td>27.54 (+215%)</td><td>27.58 (+216%)</td><td>30.61 (+250%)</td><td>NR</td><td>NR</td><td>NR</td><td>8.15 (-21%)</td><td>8.15 (-21%)</td><td>5.31 (-48%)</td><td>64.32 (-21%)</td><td>64.27 (-21%)</td><td>64.08 (-21%)</td></tr>
<tr><td>Ottawa-Montreal</td><td>14.96 (+65%)</td><td>16.16 (+78%)</td><td>17.89 (+97%)</td><td>4.52 (-6%)</td><td>3.59 (-26%)</td><td>4.51 (-7%)</td><td>71.85 (-6%)</td><td>71.61 (-7%)</td><td>68.95 (-10%)</td><td>8.67 (-6%)</td><td>8.64 (-7%)</td><td>8.65 (-7%)</td></tr>
</table>

Note: NR indicates not relevant, due to zero bus share in base case.

logit models. The differences imply substantially greater sensitivity of either or both of the car and bus modes to improvements in rail service. Example predictions of changes in mode probabilities for representative travelers indicate that the adoption of either of the nested logit models would result in substantially higher rail probabilities at the individual level and rail shares at the aggregate level. This result demonstrates the importance of considering alternatives to the multinomial logit structure in intercity mode choice modeling.

Differences between the nested logit models in their behavior implications and predictions raise serious questions about which of the models to adopt. Different choices result in different rail ridership estimates and different estimates of the mode source of the increased ridership. Despite the statistical rejection of the multinomial logit model and the improvement in goodness of fit, these results do not provide a satisfactory conclusion to the search for improved specification of intercity mode choice models. The apparent higher degree of sensitivity both between rail and bus and between rail and car cannot be accommodated in the nested logit structure except by including car and bus in the same nest, a choice that is inconsistent with the empirical analysis. There appears to be a need to consider more sophisticated model structures to adequately represent the substitution characteristics among these alternatives.

It is interesting to observe that these estimation results do not support the notion that improved rail service will attract a larger share of travelers from air than from other modes. However, this result is likely to represent only incremental changes in rail service. It seems reasonable to speculate that large improvements in rail service (implementation of high-speed rail or magnetic levitation) may change the competitive structure among intercity travel modes. In this case, the structure of the model may require adjustment to account for the differences in intermodal sensitivity. These results demonstrate a continuing need to develop improved intercity travel demand models.

ACKNOWLEDGMENTS

The primary author was supported during the early part of this research by the Transportation Center at Northwestern University and by a National Science Foundation Graduate Fellowship for a large part of the period. The authors are indebted to the Ontario-Quebec Rapid Train Task Force for allowing the use of the data and to Bruce T. Williams of KPMG Peat Marwick for assembly and documentation of the estimation data file.

REFERENCES

1. *Special Report 233: In Pursuit of Speed: New Options for Intercity Passenger Transport.* TRB, National Research Council, Washington, D.C., 1991.
2. *Terminal Area Forecasts 1987–2000.* FAA, U.S. Department of Transportation, 1987.
3. *Future Development of the U.S. Airport Network: Preliminary Report and Recommended Study Plan.* TRB, National Research Council, Washington, D.C. 1988.
4. Daley Insists Third Airport Dead, Buried. *Chicago Tribune*, July 2, 1992, Sec. 1, p. 1.
5. Searles, D. Critics Haven't Grounded Denver Airport. *Chicago Tribune*, July 19, 1992, Sec. 1, p. 1
6. Moon, F. B. Houston Moves Forward Again. *Civil Engineering*, Vol. 61, No. 2, 1991, pp. 40–43.
7. Hartje, R. L., and G. S. Pfeffer. Private Toll Road: SR 91 Toll Express Lanes. *Compendium of Technical Papers.* ITE, Washington, D.C., 1991.
8. Dees, D. C., and J. W. Guyton. The Chicago-Kansas City Tollway Feasibility Study: An Investigation of Private Financing Potential. *Compendium of Technical Papers.* Institute of Transportation Engineers, Washington, D.C., 1990.
9. Baer, H., W. Testa, D. Vandenbrink, and B. Williams. *High Speed Rail in the Midwest: An Economic Analysis.* Federal Reserve Board of Chicago, Ill., 1984.
10. *Detroit-Chicago Corridor High Speed Rail Technical Report.* Intercity Transportation Planning Division, Bureau of Transportation Planning, Michigan Department of Transportation, Lansing, 1985.
11. *Preliminary Implementation Plan.* National Maglev Initiative, U.S. Army Corps of Engineers, June 1990.
12. KPMG Peat Marwick and F. S. Koppelman. *Analysis of the Market Demand for High Speed Rail in the Quebec-Ontario Corridor.* Report produced for Ontario/Quebec Rapid Train Task Force. KPMG Peat Marwick, Vienna, Va., 1990.
13. Peat Marwick Main and Co. and F. S. Koppelman. *Preparation of Base Year Service Characteristics and Development of Travel Demand Models: Working Paper #3.* Report submitted to Ohio High Speed Rail Authority, June 1989.
14. Westin, R. B., and C. F. Manski. Theoretical and Conceptual Developments in Demand Modeling. In *Behavioural Travel Modeling* (D. A. Hensher and P. R. Stopher, eds.), Croom Helm, London, 1979.
15. Hensher, D. A. Efficient Estimation of Hierarchical Logit Mode Choice Models. *Proceedings of the Japanese Society of Civil Engineers*, Vol. 425, No. 4, 1991, pp. 17–28.
16. Stephanedes, Y. S., V. Kummer, and B. Padmanabhan. A Fully Disaggregate Mode-Choice Model for Business Intercity Travel. *Transportation Planning and Technology*, Vol. 9, No. 1, 1984.
17. Stopher, P. R., and J. N. Prashker. Intercity Passenger Forecasting: The Use of Current Travel Forecasting Procedures. *Transportation Research Forum Proceedings: 17th Annual Meeting*, Vol. XVII, No. 1, 1976, pp. 67–75.
18. Grayson, A. Disaggregate Model of Mode Choice in Intercity Travel. In *Transportation Research Record 835,* TRB, National Research Council, Washington, D.C., 1981, pp. 36–42.
19. Ellis, R. H., P. R. Rassam, and J. C. Bennett. Consideration of Inter-Modal Competition in the Forecasting of National Intercity Travel. In *Highway Research Record 369*, HRB, National Research Council, Washington, D.C., 1971, pp. 253–261.
20. Ortuzar, J. de D. Nested Logit Models for Mixed-Mode Travel in Urban Corridors. *Transportation Research A*, Vol. 17, No. 4, 1983, pp. 283–299.
21. Daly, A. Estimating "Tree" Logit Models. *Transportation Research B*, Vol. 21, No. 4, 1987, pp. 251–267.
22. Sobel, K. L. Travel Demand Forecasting by Using the Nested Multinomial Logit Model. In *Transportation Research Record 775*, TRB, National Research Council, Washington, D.C., 1980, pp. 48–55.
23. Brownstone, D., and K. A. Small. Efficient Estimation of Nested Logit Models. *Journal of Business and Economic Statistics*, Vol. 7, No. 1, 1989, pp. 67–74.
24. Morrison, S. A., and C. Winston. An Econometric Analysis of the Demand for Intercity Passenger Transportation. In *Research in Transportation Economics*, Vol. 2 (T. E. Keeler, ed.), JAI Press, Inc., Greenwich, Conn., 1983, pp. 213–238.
25. Ben-Akiva, M. E., and S. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand.* Massachusetts Institute of Technology, Cambridge, Mass., 1985.
26. McFadden, D. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics* (P. Zarembka, ed.), Academic Press, New York, N.Y., 1973.
27. Hensher, D. A. Sequential and Full-Information Maximum Likelihood Estimation of a Nested Logit Model. *The Review of Economics and Statistics,* Vol. 68, No. 4, 1986, pp. 657–667.

28. Cosslett, S. R. Alternative Estimation of Discrete Choice Models. In *Structural Analysis of Discrete Data with Econometric Applications*, (C. F. Manski and D. McFadden, eds.), MIT Press, Cambridge, Mass., 1981.
29. Forinash, C. V. A Comparison of Model Structures for Intercity Travel Mode Choice. Master's thesis (unpublished). Department of Civil Engineering, Northwestern University, Evanston, Ill., 1992.
30. Koppelman, F. S., and C. G. Wilmot. Transferability Analysis of Disaggregate Choice Models. In *Transportation Research Record 895*, TRB, National Research Council, Washington, D.C., 1982.
31. Koppelman, F. S., and C. G. Wilmot. The Effect of Omission of Variables on Choice Model Transferability. *Transportation Research B,* Vol. 20, No. 3, 1986.
32. Koppelman, F. S. Predicting Transit Ridership in Response to Transit Service Changes. *Journal of Transportation Engineering*, Vol. 109, No. TE4, July 1983.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.