Travel Demand Forecasting by Using the Nested Multinomial Logit Model

KENNETH L. SOBEL

A considerable amount of recent travel demand research has highlighted the limiting assumptions of the multinomial logit model, particularly its property of being independent of irrelevant alternatives. Nevertheless, because of its tractability, the multinomial logit formulation is likely to remain the most important of disaggregate-analysis techniques. This paper points out that, although the axiom of the independence of irrelevant alternatives is a property of the simple multinomial logit model, a generalization of that model has been developed that is virtually free from that limitation, has been shown to be effective and economically usable in practical studies, and provides a simple diagnostic capability for assessing the validity of the assumption of independence of irrelevant alternatives in any given situation. The generalized logit model—referred to here as the nested multinomial logit model—has been reported in the literature for several years, but awareness of its properties and even existence seems to be very slight. This paper provides a background on the development of the nested multinomial logit model, presents its structure (with guidelines for its use), and reports on current research that uses the nested formulation as the analysis tool.

In recent years, disaggregate approaches to analyzing travel demand have exhibited very promising characteristics, and a wide variety of advances have been achieved. Models have been developed to study not only the traditional problems of mode choice for work trips (1), but also the full range of travel decisions: frequency, time of day, destination, etc. (2-5). There are two primary rationales for the disaggregate approach--efficiency of data requirements and validity of results. Unlike models that rely on zonal averages, disaggregate models do not require analysts to discard the majority of the information that describes the distribution of important variables prior to the statistical estimation of model parameters (6). This enables development of models that exhibit high statistical validity by using only a small portion of the data otherwise necessary. Second, because travel decisions and factors that influence travel decisions are measured and analyzed at the level of the household or the individual, it seems more plausible that actual behavioral relationships may be reflected in successful models rather than in the simple exploitation of ecological correlations in the data (2). This provides increased confidence in forecasts (which of course requires some degree of faith in the behavioral representations of the models).

Most disaggregate models have been formulated from the concept of random utility, which assumes that individuals' evaluation of available alternatives and their attributes can be conceptually described by utility functions and that the choice process can be conceptually described as the selection of the alternative that has the greatest utility (7). However, it has also been explicitly recognized that all the important components of utility functions cannot be observed or measured, so that in practice the utility functions (U) of alternative i are typically represented by a deterministic portion (V) and a (usually additive) random portion (ϵ):

$$U_i = V_i + \epsilon_i \quad (1)$$

The deterministic portion of the utility function is composed of the observable characteristics of the alternatives and the decision maker and measures the average (systematic) tastes of decision makers within each category of socioeconomic descriptors. The random portion of the utility function contains the unobservable attributes of the alternatives and the decision maker, which includes idiosyncratic variations in taste that may be present in the population of decision makers (also called random taste variation).

The popular multinomial logit (MNL) model is based on the assumptions that the random components of the utility function are independently and identically distributed by means of the negative reciprocal exponential distribution:

$$\text{Prob}(\epsilon_i \leq c) = \exp[-\exp(-c)] \quad (2)$$

This is equivalent to assuming that random taste variation within a population of interest does not exist and that the effects of unobservable attributes of individuals and alternatives are uncorrelated across individuals or alternatives. Specifying the random components of the utility in this fashion allows for the derivation of the simple MNL model (8):

$$P_i = \exp(V_i)/\sum_{i}^{A} \exp(V_i) \quad (3)$$

where P_i is the probability that a decision maker will choose alternative i from the set of A possibilities and V_i is the deterministic portion of the utility function of alternative i.

In recent years, the assumptions of the MNL model (lack of random taste variation and uncorrelated disturbance terms across alternatives and individuals) have been criticized as being overly restrictive and, in some cases, blatantly counter to observed behavior. This has been especially true insofar as these assumptions have led to the notorious property of independence of irrelevant alternatives (IIA) (9-11). The IIA property states that the relative odds that an individual will select one alternative from an available pair of alternatives is independent of the presence or absence of any other alternatives. Although this property may be quite reasonable in many cases and in fact is useful for the prediction of demand for a new alternative, it is also easy to construct examples in which the IIA property yields false results.

Consider the infamous problem of the red bus versus the blue bus: A given market is initially served by two modes--automobile and a bus line with red buses. The automobile mode has two-thirds of the market, so the ratio of automobile to bus probabilities is 2:1. If blue buses are introduced into the bus line (with relevant characteristics identical to those of red buses), we would expect the new market shares to be two-thirds for automobile and one-sixth for each bus mode (red or blue). However, because of the IIA property, the MNL model will predict the automobile's new market share to be only twice that of the red bus, not four times as large. Further, because the relevant characteristics of the red and blue bus modes are identical, their market shares will be predicted to be equal. Thus, the ratio of market shares predicted by the MNL for automobile to red bus to blue bus is 2:1:1, or one-half to one-fourth to one-fourth.

With the recent development of improved methods of statistically estimating multinomial probit models (12-14), many researchers have shifted attention to that model, because of its ability to represent explicitly random taste variation within a sample population and especially because of its ability to account explicitly for covariance among the unobserved attributes of the alternatives' utility functions, thereby overcoming the IIA restriction inherent in the MNL model (10). The added flexibility of the multinomial probit model is not gained without paying a price: Its generality is derived by the estimation of many more parameters than are necesary (or possible) when the logit model is applied. For example, one empirical comparison test of equivalent logit and probit models required the estimation of 34 probit parameters and only 7 logit parameters (12). This suggests that the statistical efficiency of each of the coefficient estimates may be lower in probit models than in logit models, which yields greater standard errors of estimates or requires larger data samples (12). Of course, it is a reasonable criticism to state that, when two models require the estimation of 34 and 7 parameters, they can hardly be considered equivalent. In the case cited, however, the key to the comparison lies in testing the probit model's added flexibility--this flexibility is provided by the additional estimated coefficients. Therefore, to restrict the number of parameters to be equal for the two models would defeat the purpose of the comparison. Also, the computational requirements of estimation appear to be between 2 to 10 times as great for probit models as for logit models (10), a factor that may have practical importance in the production environment of ongoing studies. In addition, there is reported experience that the estimation properties of multinomial probit models may not be well behaved (their likelihood functions may exhibit multiple local optima), which possibly would confound attempts to solve for the the maximum-likelihood coefficient estimates in some circumstances (15).

One rationale of this paper concerns the frequent statement that the MNL model cannot account for interdependence among alternatives. In fact, this statement is only completely valid for a restricted variation of MNL models, called the simple MNL model. In addition to the simple MNL model, the more-general nested MNL logit (also called the structured or hierarchical MNL model by some authors) retains many of the desirable characteristics of simple MNL formulations but also explicitly represents many of the possible correlations of observed attributes across alternatives and does not therefore suffer from the restrictions of the IIA axiom in situations in which it is not warranted. Furthermore, the model also provides an explicit statistical diagnostic of the appropriateness of assuming independence across alternatives. Therefore, when the purpose is to transcend the limitations inherent in the IIA property of the simple MNL model, to represent interalternative correlations of the utility function's disturbance terms, or to test whether either of the above possibilities is valid, it is not necessary to abandon the advantageous computational properties of the MNL model. Instead, one can accomplish those more-general investigations with the nested MNL model. Of course, for situations in which it is desirable to represent or test for the existence of significant random taste variation, the nested MNL model will not be the appropriate analysis tool; fully generalized multinomial probit approaches will probably be required instead. However, many multinomial probit analyses have been performed that have restricted probit approaches, which themselves do not permit the measurement of random taste variation (16).

Although the nested MNL model has been presented in the literature, derived, and explored in the last few years, its properties (and even its existence) are not widely known. This is true in part because, in the United States, the nested MNL model has usually been applied to problems of representing multidimensional choice contexts, e.g., separate nests for mode and destination choice (4,5), even though the IIA property can also create problems within the context of a single choice dimension (several destinations may be perceived as similar to each other by travelers). One purpose of this paper is to add to the dissemination of knowledge about the nested MNL model so that unnecessary sacrifices of mathematical convenience and tractability can be avoided.

NESTED MNL MODEL STRUCTURE

To best present the nested MNL structure, it is useful to first reexamine the simple MNL model to highlight their differences. Figure 1 illustrates the model portrayed by Equation 3, in which A, the number of alternatives, equals 3. For purposes of exposition, the alternatives have been identified as bus, train, and automobile. Conceptually, each alternative is evaluated by individuals according to utility functions U_b, U_t, and U_a; furthermore, individuals are conceptualized as selecting the alternative that has the greatest value of utility. However, since the U_i's cannot be completely observed, they are written as in Equation 1. Given suitable assumptions about the distribution of the ε_i's, Equation 3 is derived.

If there are reasons to believe that the alternatives are not completely independent, one can postulate that a particular nested structure applies or, alternatively, one can test the validity of all possible nested structures as well as the simple (MNL) structure. Figure 2 shows one nested structure that seems to be a likely candidate for

Figure 1. Simple MNL model.

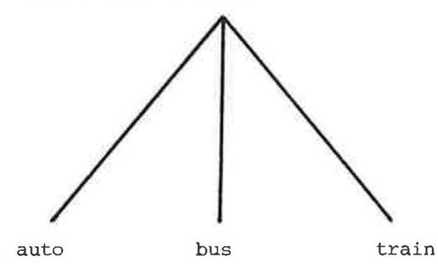
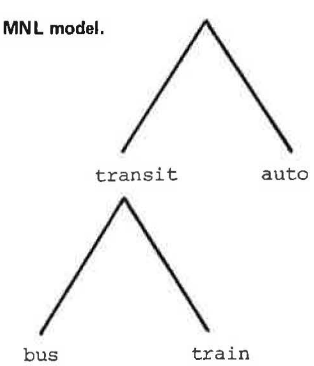


Figure 2. Nested MNL model.



testing. In this situation, each individual is again conceptually assumed to evaluate each of the alternatives that has the same utility function as specified by the simple MNL model. However, there is also a composite utility of the nest, which in this case represents public transit. The composite utility includes the expected value of the maximum utility of the members of the nest, given by

$$I_{b,t} = E[\max(U_i)] = n \sum_{i1}^{N} \exp(V_i) \quad (4)$$

where $I_{b,t}$ is the expected maximum utility of the members of the public-transit nest, N is the number of available alternatives in the nest (N < A), and all other symbols are as defined previously.

The nest's composite utility is then written as

$$V_{b,t} = \theta I_{b,t} + \underline{g} W_{b,t} \quad (5)$$

where θ is an estimated coefficient, $\underline{g}$ is a vector of estimated coefficients, and $W_{b,t}$ is a vector of attributes common to all members of the nest.

The nested MNL model shown in Figure 2 can be estimated by using standard logit estimation software in two stages: First a simple binary logit model between bus and train is estimated; the results allow the calculation of the expected maximum utility of the nest $I_{b,t}$ according to Equation 4. This value is then entered as a typical independent variable that has the $\underline{W}_{b,t}$ variables and the characteristics of automobile into a second-level simple binary logit model between the public-transit nest and automobile.

For prediction, the first-level logit model yields P(b|b,t) and P(t|b,t), the conditional probabilities of the bus or train given that the choice is constrained to public transit. The second-level logit model yields P(b,t) and P(a), the marginal probabilities of public transit and automobile, respectively. To calculate bus and train choice probabilities, Equations 6 and 7 are invoked:

$$P_b = P(b|b,t) \cdot P(b,t) \quad (6)$$

$$P_t = P(t|b,t) \cdot P(b,t) \quad (7)$$

The automobile choice probability (P_a) is given directly by the second logit model.

A critically important feature of the model concerns acceptable values of θ, the coefficient of the expected maximum utility of the nest. It can be proved [see the report by Williams and Ortuzar (17)] that θ must satisfy $0 < \theta \leq 1$ and that, if $\theta \leq 0$ or $\theta > 1$, pathological forecasts may result. If $\theta < 0$, then improving the utility of one member of a nest (say, V_b) can decrease the choice probability of selection P_b of that alternative. If $\theta = 0$, then an improvement in the utility of one or both members of a nest will not change the choice probability of the nest. If $\theta > 1$, then improving the utility of one member of a nest (say, V_b) will not only improve its choice probability P_b but may also improve the choice probability of other members of the nest (here, P_t). If $\theta = 1$, then the choice-probability calculations yield algebraically equivalent results to those of the simple MNL model.

The concept of separable logit models linked by measures of inclusive utility is not new. Even the particular formulation of Equation 4 as the functional form of the linking measure was tested as early as 1973 (2). However, in the early tests, the consistency of Equation 4 with the underlying utility maximization theory was not recognized. This is shown by the selection of other composition rules or by the rejection of any composition rule as unfounded (2,4,18,19). Soon afterward, however, the behavioral consistency of the composition rule embodied by Equation 4 was formally derived and proved by several researchers almost simultaneously (20-23).

NESTED MNL MODEL ISSUES

Structural Alternatives and Diagnosis

Use of the nested MNL model results in a new degree of freedom in the problem of specifying a model. Not only must the analyst (a) specify the functional form of the choice probabilities (logit, probit, etc.), (b) identify the available choice set for the members of the relevant population, (c) select the appropriate set of explanatory variables, and (d) define the functional form of the utility functions, but also he or she must decide on or test the structure of the model a priori. Figure 3 displays a number of feasible structures for the cases of two, three, and four fundamental choice alternatives (a-d). Clearly the number of structural alternatives increases much faster than the number of choice alternatives. Furthermore, the selected structure may interact with the desirable variable specification, so that, when a satisfactory set of variables is tested in the context of one structure, it may prove to be unsatisfactory when imbedded in another structure. This, of course, would considerably increase the complexity of searching for the best model for a given choice context.

As described earlier in this paper, there is an important restriction on the values that the coefficients of the expected maximum utilities (the θ's) can take. Specifically, θ must satisfy $0 < \theta_i \leq 1$, where θ_i represents the coefficient of the expected maximum utility of the ith-level nest. Furthermore, if $\theta_i = 1$, then the linked nest at level i is mathematically equivalent to the simple MNL model at that level. (As an example, referring to the four-alternative case of Figure 3, if the θ that corresponds to the expected maximum utility $I_{a,b,c}$ of structure 21 is equal to 1, then structure 21 is mathematically equivalent to structure 11.) Clearly, structure 1 (the simple MNL structure) is the special case of all other structures when all possible θ's have values of 1.

These properties suggest a technique by which an analyst can statistically test whether particular structures can be rejected and whether the IIA property is appropriate for the situation that is being examined. Each feasible structure (after prescreening to eliminate theoretically unreasonable structures) can be estimated in turn. The tested structure is rejected if θ does not satisfy the constraint $0 < \theta \leq 1$, and, if θ is not very different from 1, its nest and structure can be evolved into a less-general form. If all θ's equal 1, then the IIA property cannot be rejected (alternatives cannot be empirically indicated to be interdependent) and the simple MNL model is likely to be appropriate.

When an estimate for θ results in a value of approximately 1, it is preferable to reestimate the model without the separate nest. Although the mathematics of a nested MNL model with $\theta = 1$ is equivalent to a simple MNL model, the statistical results of the two formulations may not be identical for three reasons. First, the values of the coefficients of the lower-level logit model (which are used to calculate the value of the nest's expected maximum utility I) are not known with certainty. Their errors create an additional source

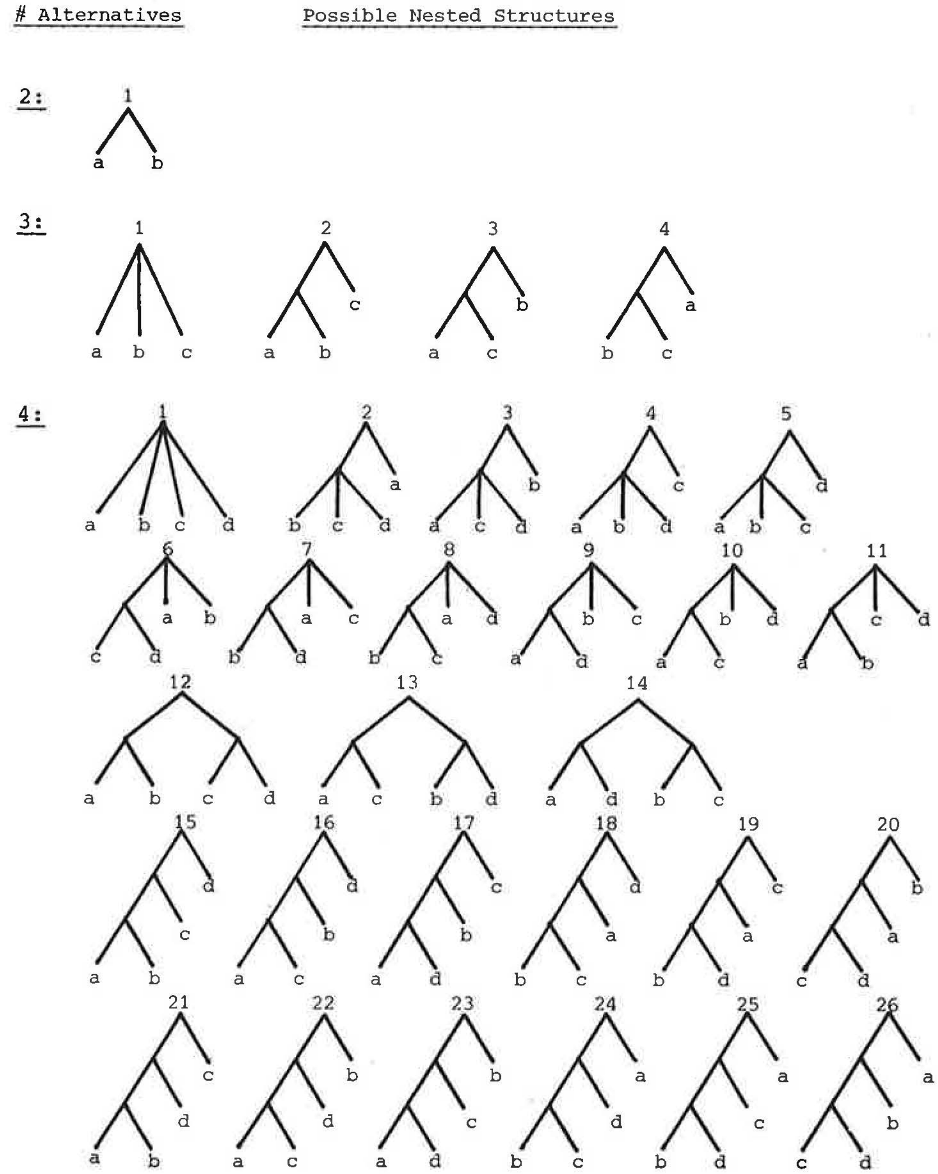


Figure 3. Nested MNL structural alternatives.

of measurement error in the value of I; this measurement error affects the estimated coefficients of the higher-level logit model. This problem would be eliminated if all the nested MNL coefficients were estimated in one step. Second, since the same amount of data is used to calibrate either the simple or the nested MNL model, the estimates of the coefficients of the simple MNL model will be statistically more efficient since there is one less coefficient that requires estimation (there is no θ). Third, only a subset of the full data set is used to estimate the coefficients of the utility functions of the members of the (lower-level) nest of the nested MNL model, although the complete data set is used for estimating all coefficients of the simple MNL model. This more-complete use of data results in statistically better coefficient estimates.

The computational advantage of nested logit estimation when compared with probit estimation loses its importance when the process of structural testing is considered. Although any given probit estimation may require 2-10 times the computational resources of logit estimation (10), the probit results show the degrees of interdependence between all possible pairs of alternatives. In contrast, a single nested MNL model estimation measures only as many sets of interdependencies as there are θ's in the model; many nested MNL model estimations may be required to yield most of the information that results from one (albeit complex) probit estimation.

Goodness-of-Fit Measures

Goodness-of-fit measures for logit models depend on the values of the logarithm of the models' likelihood function when the coefficients assume various values. In general, the value of the likelihood function (L) is given by

$$L = \prod_{ij} P_{ij}^{N_{ij}} \qquad (8)$$

where P_{ij} is the probability that j would choose alternative i, and N_{ij} equals 1 if individual j was observed to choose alternative i, 0 otherwise. P_{ij} is found from Equation 3. The logarithm of L is usually denoted L*. The value of L* when all the coefficients in V are set to zero is written L*(0) and represents the maximum amount of uncertainty that can be removed by developing a perfect model; $L^*(0)$ corresponds to an initial state of information that all alternatives are equally likely. Because of the way in which it is defined, $L^*(0)$ is a large negative number. When the coefficients in V_i are set to their maximum-likelihood estimated values, the result is $L^*(\beta)$, a smaller negative number [a value of 0 for $L^*(\beta)$ would indicate a perfect model]. $L^*(\beta)$ corresponds to a final state of information about the likelihood of alternatives when the information in V is fully known. Most logit estimation software packages routinely report both $L^*(0)$ and $L^*(\beta)$ as part of their output. When all coefficients in V are set to zero except the coefficients of a full set of alternative-specific constant terms, the result is $L^*(C)$, a negative number that lies within the range $L^*(0) \leq L^*(C) \leq L^*(\beta)$. $L^*(C)$ corresponds to a second initial state of information that alternatives are as likely to be chosen by any individual as are their aggregate market shares.

The value of $L^*(C)$ can also sometimes be calculated without the estimation of a restricted model. A formula for $L^*(C)$ for a binary model has already been reported (24), and the following equation generalizes that result for a model among N alternatives in which all individuals have all N alternatives available to them or in which the unavailability of alternatives is independent across individuals:

$$L^*(C) = \sum_{i=1}^{N} x_i [\ln(X_i/Y_i)] \quad (9)$$

where X_i equals the number of observations in the estimation data set that have selected alternative i, and Y_i equals the total number of observations in the estimation data set that had alternative i available (including those that selected alternative i).

Because the dependent-variable observations of logit models are discrete, or qualitative (e.g., bus, automobile), a coefficient of determination (R^2) cannot be calculated as is done with regression analysis. Statistics similar to R^2 are constructed from the values of L^* given above and are called ρ^2 (8,24). In particular,

$$\rho^2 = 1 - [L^*(\beta)/L^*(0)] \quad (10)$$

$$\rho_c^2 = 1 - [L^*(\beta)/L^*(C)] \quad (11)$$

Both lie between 0 and 1, although the corrected (ρ_C^2) allows comparisons between models estimated with observation sets that have different market shares.

At first glance, the development of overall measures of goodness of fit for linked sequential estimated logit models appears to be a complicated task. In reality, measures equivalent to ρ^2 suggested by McFadden (8) and the ρ_C^2 suggested by Tardiff (24) can easily be constructed. The corresponding equations are as follows:

$$\rho^2 = 1 - \{[L_1^*(\beta) + L_2^*(\beta) + \ldots + L_j^*(\beta)] / [L_1^*(0) + L_2^*(0) + \ldots + L_j^*(0)]\} \quad (12)$$

$$\rho_c^2 = 1 - \{[L_1^*(\beta) + L_2^*(\beta) + \ldots + L_j^*(\beta)] / [L_1^*(C) + L_2^*(C) + \ldots + L_j^*(C)]\} \quad (13)$$

where the subscripts 1 through j refer to the j simple MNL models in the structure of interest.

Simultaneous or Sequential Choice

Clearly, as described in this paper and elsewhere, the estimation process that uses nested MNL models has so far been sequential: Lower-level choices are estimated first, then inclusive utilities of nests are calculated, and last upper-level choices are estimated. (These estimations could be done simultaneously, and more is mentioned on this issue later.) The forecasting process that uses nested MNL models is also sequential, although the direction of sequence is less clear. First, lower-level models are applied to calculate conditional choice probabilities and inclusive utilities (moving up the tree); then marginal choice probabilities and trip volumes are calculated (moving down the tree).

These sequences notwithstanding, the fundamental question whether nested MNL models imply a particular sequence of individual decisions may not be meaningful. When there is a clear reason to presuppose a particular sequence (say, one nest is the mode choice for shopping trips and the higher nest represents residential location), then the nested MNL can be used to represent a choice sequence (3,5). On the other hand, if the nests are intended to represent varying degrees of closeness among alternatives (one nest represents access mode and the higher nest models line-haul mode), the nested MNL can clearly be interpreted as a model of simultaneous choice broken into steps merely for reasons of computational convenience (21).

PLANNING APPLICATION

As part of a larger currently ongoing regional-planning study of the Rotterdam-Hague metropolitan area (25), mode-choice models are being developed to represent travelers' decisions among the six fundamental alternatives: automobile driver (D), automobile passenger (P), public transit (T), walk (W), bicycle (B), and moped (M). Among the preliminary mode-choice models estimated for travel to work, there were two models that had identical specifications in all respects except for structure. Figure 4 illustrates the two nested structures: (a) the four-alternative structure A, made up of the combined D and P alternatives and the slow modes (the W, B, and M alternatives), and (b) the three-alternative structure B [automobile (D and P), T, and the slow modes]. The approximate split of travel in the study area among these modes is D, 22 percent; P, 8 percent; T, 6 percent; W, 33 percent; B, 28 percent; and M, 3 percent; or 30 percent for automobile, 6 percent for transit, and 64 percent for the slow modes.

The four-alternative model (structure A) shown in Figure 4, which was based on 726 observations, had an overall ρ_C^2 of 0.321. The coefficient of the slow-mode expected maximum utility was 0.413; the t-ratio was 1.44. (Note that, when calculated by many of the standard MNL estimation software packages, t-ratios at upper levels of a nested structure are biased upward. Examples cited in this paper have not been corrected for such bias.) The slow-mode expected maximum utility variable was calculated from a lower-level submode-choice model of work travel that had 21 variables (Table 1) and the upper-level main mode-choice model included 33 other variables as well (Table 2).

Structure A in Figure 4 was converted to structure B by estimating a second lower-level submode-choice model between the automobile alternatives (D or P). The nine variables used in structure A and associated with the D and P alternatives formed the specification of the automobile submode-choice model (Table 3), which was used to compute the expected maximum automobile utility variables for the main mode-choice model (Table 4). The other

Figure 4. Nested mode-choice structures.

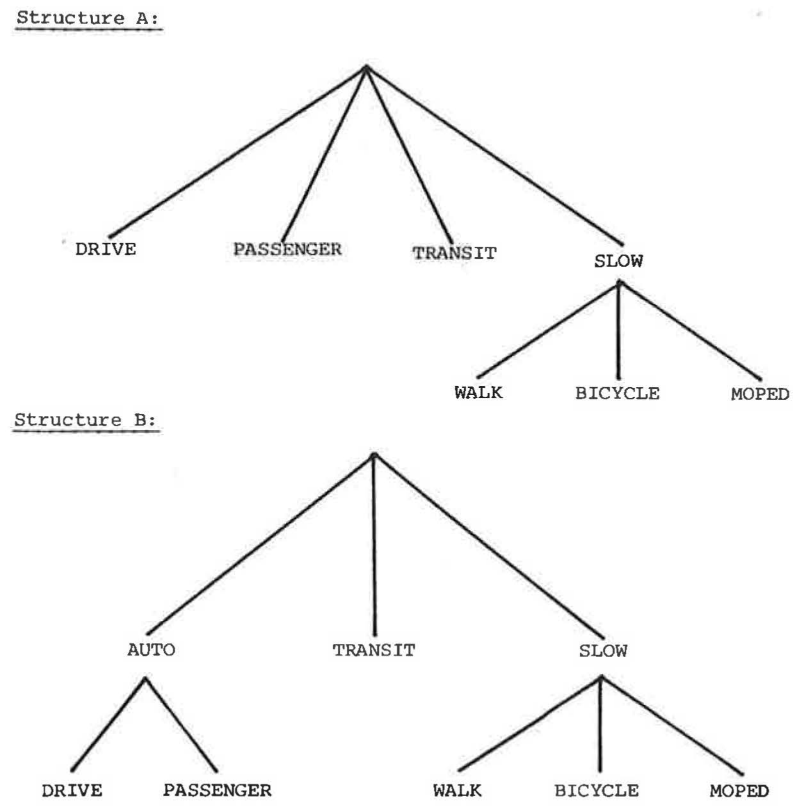


Table 1. Submode-choice model for slow modes.

Variable	Coefficient Value	t-Ratio
W-constant (=1)	1.25	2.53
W-distance		
0-1 km	-1.74	-3.32
>1 km	-0.558	-4.60
W-1		
If age <25 years	-0.549	-2.07
If lunch-time trip	-0.184	-0.810
If income >$10 500	-1.62	-6.20
M-constant (=1)	-3.86	-9.71
M-distance		
0-9 km	0.273	5.16
>9 km	0.0876	1.11
M-1 if age		
<20 years	1.17	3.74
25-45 years	-0.626	-2.03
M-1 if departure before 8:00 a.m.	0.818	2.87
W-1 if unfixed destination	0.333	0.911
M-1		
If unfixed destination	-0.390	-0.696
If white collar	-0.472	-1.37
W-1 if blue collar	-0.991	-3.13
M-1		
If blue collar	0.467	1.56
If no driver's license	1.01	3.41
M,W-1 if part-time or commercial worker	0.665	2.91
W-1 if service worker	0.436	1.23
W-population density, origin	0.000 082 1	3.28

25 variables previously used in structure A were left unchanged in the three-alternative model (structure B), which was then estimated with 765 observations of work-travel modes. The overall ρ_c^2 for structure B was 0.333. The coefficient of the slow-mode expected maximum utility was 0.384; the t-ratio was 1.36. The coefficient of the automobile-mode expected maximum utility was 0.477; the t-ratio was 2.7.

Table 2. Structure-A main mode-choice model.

Variable	Coefficient Value	t-Ratio
D-constant (=1)	-1.34	-2.05
T-constant (=1)	0.150	0.243
P-constant (=1)	-1.12	-1.89
S-expected maximum utility	0.413	1.44
D,P,T-in-vehicle time	-0.005 17	-0.551
D,P,T-cost	-0.0875	-0.945
T-final walk time	-0.003 94	-0.401
T-outbound headway	-0.0481	-2.97
T-return headway	-0.109	-3.90
T-1 if Rotterdam destination	0.694	1.94
S-distance		
0-10 km	-0.355	-7.32
10-25 km	0.315	3.62
>25 km	-0.742	-1.85
S-1		
If departure after 5:30 p.m.	-0.376	-1.13
If lunch-time trip	0.395	
If age <30 years	0.304	1.28
If male	0.920	2.67
If Hague destination	0.940	2.71
D-number of cars	0.812	2.36
D-1		
If male	1.59	4.06
If parking cost >0 and arrival after 9:00 a.m.	1.23	1.05
P-distance	0.004 76	1.29
P-parking cost	0.168	0.940
P-household density, origin	-0.000 194	-1.28
P-employment density, destination	-0.000 066 9	-0.532
P-1 if unfixed destination	0.103	0.212
D,P-1		
If male	-0.220	-0.442
If age >55 years	-1.05	-2.78
D,P-number of cars per license		
If 1 car	0.728	1.75
If 2+ cars	0.261	0.639
D,P-1		
If parking cost >0 and arrival after 9:00 a.m.	-1.14	-0.926
If no driver's license	-2.32	-6.07
If white collar	-0.508	-2.25
If peak-period trip	-0.866	-2.84

Clearly, the value of θ_A in structure B is sufficiently and significantly unequal to 1.0 to indicate the lack of independence between the D and P modes, which helps to explain the improved summary statistics of the structure-B model despite its use of a seemingly lower number of variables.

For Tables 1-4, it should again be stressed that these were preliminary models, already superseded by revisions typically necessary in the course of an ongoing study (25). The modal symbols preceding each variable description (B-, W-, M-, T-, etc.) show the alternative (i.e., utility function) with which that variable is associated. Table 5 summarizes the statistics for each of the models compared.

FUTURE RESEARCH

Three areas for future research will be mentioned; they are areas likely to yield high payoffs or quick results and insights or both. In reality, the nested MNL model is simply a special case of a class of generalized extreme value (GEV) models (22). However, to my knowledge, it is the least-restrictive form of GEV model implemented in an operational sense thus far. Other GEV models should be pursued through at least the proof-of-concept stage, especially insofar as they may be made to represent random taste variation within a mathematically convenient framework.

Futher investigation into the numerical solution of probit models should be pursued with two goals: to learn about the potentially pathological behavior of the probit's likelihood function (15) and to reduce further the computational burden associated with evaluating the likelihood function. Research into the comparison of probit and logit models [for example, the report by Horowitz (10)] should be expanded to consider nested MNL models so as to draw more-meaningful conclusions.

Finally, accessible and user-oriented software should be developed to allow for the simultaneous estimation of all levels of coefficients of a nested MNL model. Although this presents no new theoretical problems, the computer programming may be quite complex. Nevertheless, the consequence of

Table 3. Submode-choice model for automobile models.

Variable	Coefficient Value	t-Ratio
D-constant (=1)	0.0995	0.177
D-number of cars	1.29	3.32
D-1		
If male	0.930	2.78
If parking cost >0 and arrival after 9:00 a.m.	2.65	2.13
P-distance	0.009 31	4.01
P-parking cost	0.390	2.99
P-household density, origin	-0.000 413	-2.75
P-employment density, destination	-0.000 251	-1.89
P-1 if unfixed destination	-0.653	-1.60

Table 4. Structure-B main mode-choice model.

Variable	Coefficient Value	t-Ratio
A-constant (=1)	-1.61	-2.94
T-constant (=1)	-0.222	-0.376
A-expected maximum utility	0.477	2.70
S-expected maximum utility	0.384	1.36
A,T-in-vehicle time	-0.006 94	-0.770
A,T-cost	-0.108	-1.21
T-final walk time	-0.004 77	-0.512
T-outbound headway	-0.0313	-2.28
T-return headway	-0.0977	-4.05
T-1 if Rotterdam destination	0.826	2.33
S-distance		
<10 km	-0.363	-7.66
10-25 km	0.326	3.88
>25 km	-0.780	-1.71
S-1		
If departure after 5:30 p.m.	-0.464	-1.41
If lunch-time trip	0.365	1.54
If age <30 years	0.270	1.16
If male	0.822	2.44
If Hague destination	0.885	2.59
A-1		
If male	0.346	0.870
If age >55 years	-0.987	-2.64
A-number of cars per license		
If 1 car	1.81	4.98
If 2+ cars	1.07	2.98
A-1		
If parking cost >0 and arrival after 9:00 a.m.	-1.00	-1.42
If no license	-1.97	-3.47
If white collar	-0.454	-2.05
If peak-period trip	-0.755	-2.55

Table 5. Summary statistics for models.

	Model				
Statistic	Submode Choice for Slow Modes[a]	Submode Choice for Automobile[b]	Main Mode Choice[c]	Overall Structure	Corrected Overall Structure[d]
Structure A					
L*(0)	-802.44		-868.56	-1671.00	-1241.67
L*(C)	-643.66		-740.85	-1384.51	-1041.61
L*(β)	-469.06		-470.71	-939.77	-686.83
ρ^2	0.415		0.458	0.438	0.447
ρ_m^2	0.198		0.147	0.171	0.161
ρ_c^2	0.271		0.365	0.321	0.341
Structure B					
L*(0)	-802.44	-765.93	-747.59	-2315.96	-1345 46
L*(C)	-643.66	-263.56	-711.22	-1618.44	-1078.40
L*(β)	-469.06	-231.56	-379.08	-1079.70	-659.31
ρ^2	0.415	0.697	0.493	0.534	0.510
ρ_m^2	0.198	0.656	0.049	0.301	0.198
ρ_c^2	0.271	0.121	0.467	0.333	0.389

[a] Number of observations by mode were as follows: walk, 146; bicycle, 480; moped, 107 (total = 733).
[b] Number of observations by mode: driver, 1034; passenger, 71 (total = 1105).
[c] Number of observations by mode: structure A—driver, 313; passenger, 41; transit, 89; slow, 283 (total = 726); structure B—automobile, 354; slow, 322; transit, 89 (total = 765).
[d] Adjustments made to correct for inconsistencies due to varying sample sizes.

sequential estimation is a loss of statistical efficiency, which may be severe.

ACKNOWLEDGMENT

Grateful thanks are due Andrew Daly for many hours of enlightening discussions and for a careful review of an earlier draft of this paper. Further benefits have come from talks with Moshe Ben-Akiva and correspondence with Yehuda Gur, Frank Koppelman, Juan de Dios Ortuzar, and Timothy Tardiff. James Berkovec, Douglas Bell, and Andrew Daly each played a major role in developing the sample application given here. Thanks are also due the Netherlands Ministry of Transport and Public Works for supporting some of the work represented here and for permitting preliminary results to be disseminated. Any errors in fact and all opinions are my sole responsibility and do not necessarily reflect the views of the Netherlands Ministry of Transport and Public Works or of Cambridge Systematics.

REFERENCES

1. S. L. Warner. Stochastic Choice of Mode in Urban Travel. Northwestern Univ. Press, Evanston, IL, 1962.
2. M. E. Ben-Akiva. Structure of Passenger Travel Demand Models. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Ph.D. thesis, 1973.
3. The SIGMO Study Methodology, Volume 2. Cambridge Systematics, Cambridge, MA; Netherlands Ministry of Transport and Public Works, the Hague, 1977.
4. A Disaggregated Behavioral Model of Urban Travel Demand. Charles River Associates, Cambridge, MA; U.S. Department of Transportation, 1972.
5. E. R. Ruiter and M. E. Ben-Akiva. System Structure, Component Models, and Application Procedures. TRB, Transportation Research Record 673, 1978, pp. 121-128.
6. C. R. Fleet and S. R. Robertson. Trip Generation in the Transportation Planning Process. HRB, Highway Research Record 240, 1968, pp. 11-31.
7. D. Brand. Travel Demand Forecasting: Some Foundations and a Review. In Urban Travel Demand Forecasting, HRB, Special Rept. 143, 1973, pp. 239-282.
8. D. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics (P. Zarembka, ed.), Academic Press, New York, 1974.
9. Charles River Associates. Disaggregate Travel Demand Models. NCHRP, Project 8-13, Phase 1, Vol. 2, 1976.
10. J. Horowitz. The Accuracy of the Multinomial Logit Model as an Approximation to the Multinomial Probit Model of Travel Demand. U.S. Environmental Protection Agency, 1979.
11. D. McFadden, K. Train, and W. Tye. An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model. TRB, Transportation Research Record 637, 1977, pp. 39-46.
12. R. L. Albright, S. R. Lerman, and C. F. Manski. Report on the Development of an Estimation Program for the Multinomial Probit Model. Cambridge Systematics, Cambridge, MA; U.S. Department of Transportation, 1977.
13. C. F. Daganzo, F. Bouthelier, and Y. Sheffi. Multinomial Probit and Qualitative Choice: A Computationally Efficient Algorithm. Transportation Science, Vol. 11, 1977, pp. 339-358.
14. J. A. Hausman and D. A. Wise. A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences. Econometrica, Vol. 46, 1978, pp. 403-426.
15. C. F. Daganzo. Calibration and Prediction with Random Utility Models: Some Recent Advances and Unresolved Questions. Presented at 4th International Conference on Behavioral Travel Modeling, Eibsee, Germany, 1979.
16. F. Bouthelier and C. F. Daganzo. Aggregation with Multinomial Probit and Estimation of Disaggregate Models with Aggregate Data: A New Methodological Approach. Transportation Research, Vol. 13B, 1979, pp. 133-146.
17. H. C. W. L. Williams and J. D. Ortuzar. Behavioral Travel Theories, Model Specification, and the Response Error Problem. In Transportation Models: Proc., Summer Annual Meeting, Planning and Transport Research and Computation Co., London, England, 1979, pp. 231-262.
18. M. E. Ben-Akiva. Structure of Passenger Travel Demand Models. TRB, Transportation Research Record 526, 1974, pp. 26-42.
19. T. A. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Approach. North-Holland, Amsterdam, the Netherlands, 1975.
20. M. E. Ben-Akiva and S. R. Lerman. Disaggregate Travel and Mobility Choice Models and Measures of Accessibility. Presented at 3rd International Conference on Behavioral Travel Modeling, Tanunda, Australia, 1977.
21. A. J. Daly and S. Zachary. Improved Multiple Choice Models. In Transportation Models: Proc., Summer Annual Meeting, Planning and Transport Research and Computation Co., London, England, 1976, pp. 1-17.
22. D. McFadden. Qualitative Methods for Analyzing Travel Behavior of Individuals: Some Recent Developments. Urban Travel Demand Forecasting Project, Univ. of California, Berkeley, Working Paper 7704, 1977.
23. H. C. W. L. Williams. On the Formation of Travel-Demand Models and Economic Evaluation Measures of User Benefit. Environment and Planning A, Vol. 9, 1977, pp. 285-344.
24. T. J. Tardiff. A Note on Goodness-of-Fit Statistics for Probit and Logit Models. Transportation, Vol. 5, 1976, pp. 377-388.
25. A. Daly. Zuidvleugel Study Report 1: Proposed Model Structure. Cambridge Systematics, Cambridge, MA; Netherlands Ministry of Transport and Public Works, the Hague, 1978.