Integrated Structure of Long-Distance Travel Behavior Models in Sweden

STAFFAN ALGERS

In Sweden, large investment plans are being considered for rail and road infrastructure. At the same time, changes are taking place that either directly affect ridership (such as imposing a value-added tax on transportation) or indirectly (such as deregulation of air traffic). Clearly, there is a great need to be able to analyze how changes in price and level of service influence ridership. An overview of the models involved in a new model system for long-distance trips developed for Swedish national authorities is presented. The model system consists of nested logit models, partly estimated by the use of simultaneous estimation techniques. The trip data source is a national travel study conducted in 1984–1985. The choice structure of the model system spans from choice of access and egress mode over mode and destination choice to trip generation. There are different models for business and private trip purposes. The models contain cost parameters and mode-specific time parameters. The integrated structure implies that all variables affect all choice levels. The parameter values are reported elsewhere.

In Sweden, as in many other countries, large investment plans are being considered for rail and road infrastructure. At the same time, changes are taking place that either directly affect ridership [such as imposing a value-added tax (VAT) on transportation] or indirectly (such as deregulation of air traffic and the separation of the railway company from the authority responsible for the rail infrastructure). Clearly, there is a great need to be able to analyze how changes in price and level of service influence ridership as well as expected changes in the economic activities over a forecasting period.

Forecasting of such changes has typically been based on a linked model system that includes trip generation, trip distribution, and mode choice. In 1987, the model system was updated with a mode choice model that was estimated on disaggregate data, giving a much more policy relevant mode-choice model.

It was decided to try to further use the advantages of disaggregate modeling by using it for all steps in an integrated structure. Such a project was completed in 1991, and this paper provides an overview of the models involved in the new model system.

The term "long-distance travel" is frequently used, although is not well defined. The term refers to a specific category of trips for which there are various criteria, such as trip length. Though trip length is not necessarily the most adequate criterion for modeling, it was used in the 1984–1985 National Travel Survey in Sweden. In this survey, trips longer than 100 km (one direction) were identified as long-distance trips.

Royal Institute of Technology, Department of Traffic and Transport Planning, S–100 44 Stockholm, Sweden.

ANALYSIS OF LONG-DISTANCE TRAVEL BEHAVIOR: ANALYSIS OF DISCRETE CHOICE

As indicated earlier, the purpose of the modeling effort was to produce a system of forecasting models including mode split, trip (spatial) distribution, and trip generation. In the previous analysis, mode split was analyzed using probabilistic discrete choice models, specifically the well-known logit model (*1*). This approach was adhered to also when extending the model system to trip distribution and trip generation. Discrete choice analysis has also been applied to long-distance travel in other studies, but the use of disaggregate data, as suggested by Stopher and Prashker (*2*), has been rare. Applications of disaggregate data may be found in other work (*3–6*), but, to the knowledge of the author, no study so far has excluded access and egress mode choice, main mode choice, destination choice, and frequency choice in an integrated structure.

Logit Model

The limited space of this paper allows only a brief presentation of the logit model. The reader is otherwise referred to literature (*1*). A basic assumption in discrete choice analysis is that each alternative in the choice set of a decision maker is associated with a utility and that the decision maker chooses the alternative with the highest utility. The utility is assumed to consist one part observable and one part not observable by the analyst. Thus,

$$U_i = V_i + \varepsilon_i \qquad (1)$$

where

U_i = total utility for alternative i,
V_i = observable part, and
ε_i = unobservable part.

The unobservable part is assumed to be stochastic. This means that the alternative a decision maker would actually choose cannot be predicted but an assumption on the distribution of the stochastic part will allow one to predict the probability that it could be chosen. Thus for a population of decision makers, the share of the population choosing each alternative could be predicted.

The assumption of the distribution of the stochastic part of the utility determines the functional form of the model. In the logit model case, the assumption is that it is identically and independently Gumbel distributed. (The Gumbel distribution is fairly close to the normal distribution, the latter corresponding to the so-called probit model.) This distribution assumption implies the following formula for the probability to choose a particular alternative (the multinomial logit model):

$$P_i = \frac{e^{\mu V_i}}{\sum_{j \in C} e^{\mu V_j}} \tag{2}$$

where

P_i = probability for a decision maker to choose alternative i,
μ = a scale parameter (inversely proportional to the standard deviation of the stochastic term),
V_i = observable part of the utility, and
C = choice set of the decision maker.

In practice, V_i is often assumed to be a linear function of parameters and variables. The model can then be formulated as:

$$P_i = \frac{e^{\beta' x_i}}{\sum_{j \in C} e^{\beta' x_j}} \tag{3}$$

where β is a parameter vector (to be estimated) and x_i is a vector of variables for alternative i.

Thus, the β values reflect the sensitivity of the variables included in the model. The log of the denominator—the so-called logsum—also has a useful property in that it can be interpreted as the expected maximum utility of the alternatives in the choice set.

The assumption that the stochastic terms are independently and identically distributed is, however, fairly strong. It is probable that some alternatives to some extent share the same unobserved part of the utility function. For example, two modes to the same destination will share the unobserved part of the utility of this destination. In this case, the alternatives may be structured in classes of alternatives, such as mode alternatives and destination alternatives. A structured logit model of mode and destination choice can then be formulated as follows: A graphical presentation of the structure is shown in Figure 1.

$$P(d) = \frac{e^{\gamma' y_d + \omega \ln \sum_{m'd} \exp(\beta' x_{m'd})}}{\sum_{d' \in D} e^{\gamma' y_{d'} + \omega \ln \sum_{m'd'} \exp(\beta' x_{m'd'})}} \tag{4}$$

$$P(m|d) = \frac{e^{\beta' x_{md}}}{\sum_{m' \in M_d} e^{\beta' x_{m'd}}} \tag{5}$$

where

$P(d)$ = probability to choose destination d;
y_d = vector y of independent variables (attributes) for destination d;
γ = associated parameter vector γ, to be estimated;
D = set of p destination alternatives;
ω = logsum parameter (the ratio between the standard deviations of the error terms at the mode choice level and the destination level), to be estimated;
$P(m|d)$ = probability to choose mode m, given destination d;
x_{md} = vector x of independent variables (attributes) for mode m and destination d;
β = associated parameter vector β, to be estimated; and
M_d = set of s mode choice alternatives for destination d.

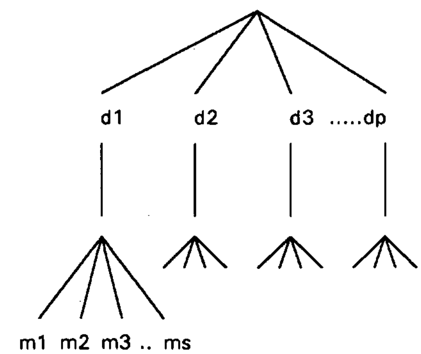

FIGURE 1 Graphical presentation of structured logit model of mode and destination choice.

The formulation of a structured model implies that the choice probabilities of the alternatives of one class is modeled conditional on the choice of the alternative of the other class. In this example, the mode choice is modeled conditional on a destination choice. Another implication is that the logsum is used to take the utilities of the alternatives of a lower class (in the sense of the graph) into account when modeling the probability for the alternatives of a higher class (or choice level).

The logsum parameter provides the connection between the choice levels and should have a value in the range of 0 to 1. If the logsum parameter takes the value of 1, then the structured model is equivalent to the normal multinomial logit model. If the value is greater than 1, unreasonable effects may be predicted, such as an increased ridership for one mode caused by an improvement of another mode (belonging to the same choice level).

Long-Distance Context

The demand for long-distance trips is thus viewed as the result of the behavior of utility-maximizing individuals, choosing among a set of mutually exclusive alternatives related to mode, destination, and trip frequency. Individuals, however, often travel together, which may influence the costs for the different modes in different ways. Therefore, effects on costs of the size and (to some extent) of the mix of persons in the traveling party were taken into account.

To define the alternatives concerning the trip, the concept of a trip must first be defined. As in other contexts, people normally start trips in their homes, visit a destination and then return to their homes. This may be called a single-destination round trip, which is how the concept of a trip was defined in the analysis. This is, of course, a simplification of the reality, as is the assumption that only one mode was used on the whole trip.

Four mode alternatives for long-distance trips were defined: car, train, air, and bus. Combined alternatives (e.g., train and air) were not defined, as the occurrence of such alternatives in the data was rare. The utility of the train, air, and bus modes may depend on the possibilities to get to and from the train or bus station and to the airport at the origin as well as at the destination. Because the access and egress modes may be of interest as policy variables and because the data permitted, the access and egress alternatives were also modeled as separate alternatives.

The destination alternatives were defined to be approximately 2,200 agglomerations and rural areas, comprising all of Sweden. Such a detailed zonal subdivision permits a more precise calculation of trip times and costs, but raises also the problem of handling many alternatives.

The frequency alternatives were defined to consist of two alternatives, to make a trip during the analyzed period or not. The fraction having made more than one trip was small.

Most variables in the analysis may be grouped into three main classes: (a) time and cost variables relating to the access and egress and main modes, (b) size variables relating to destinations, and (c) socioeconomic variables relating to the travelers.

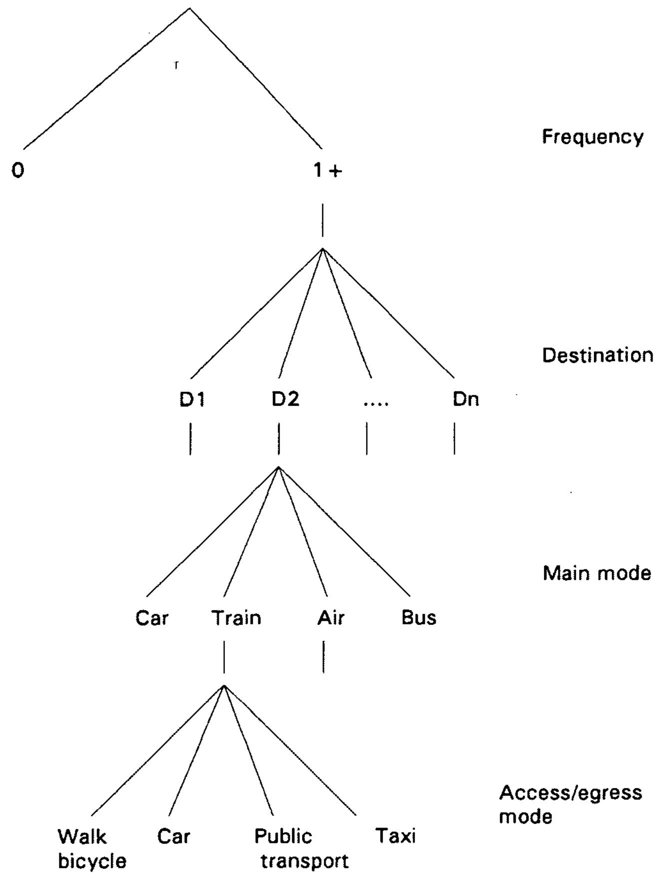

FIGURE 2 General model structure.

STRUCTURE OF LOGIT MODEL FOR LONG-DISTANCE TRAVEL BEHAVIOR

The general structure of the model is shown in Figure 2. The choice of access and egress modes is positioned at the bottom of the model. The actual structure is somewhat simplified in the figure in that the choices of access and egress modes are treated as two independent choices. At the next level is the choice of the main mode, which is influenced by the accessibility to the airport or station given by the logsum variable from the access and egress level. This variable represents the maximum expected utility from the alternatives at that level.

Destination choice comes next, being influenced by the logsum variable from the main mode level (also including the logsum variable from the access and egress level). Finally, frequency choice is positioned at the top of the structure. Frequency choice is also influenced by the logsum variable from the level below, representing the maximum expected utility from the destination alternatives (including the logsum variable from the level below). The entire structure is thus internally linked by the logsum variables, which means that changes at the lower levels will affect the higher levels.

As an example, an improvement of a bus service to an airport will, of course, cause some persons to switch from other modes to this airport (e.g., car). It will, however, also cause some persons to switch from other modes for their main trip to air because it is now easier to access the airport. A further effect is that destinations that are well served by air can now be more easily reached (because the airport is more accessible), which will cause a shift in travel to these destinations from other destinations. Finally, because accessibility is generally improved, trip frequency will also increase. The improvement of the bus service will thus influence all choices in the structure.

The magnitude of the effects will, of course, depend on the sensitivity of the model to the variables that are affected by the project under consideration. This sensitivity is embedded in the parameters of the model, which have been estimated using statistical software.

Trip Purpose

There are many reasons to expect that the sensitivity of different variables may vary by trip purpose. In this case, it was decided to estimate separate models for business trips and private trips.

Estimation

Estimating a model of this type involves some specific problems. One problem is related to the fact that the total number of alternatives in the model will be high, making it cumbersome to estimate. In this case, a stratification procedure was used, leading to 22 destination alternatives (that vary between the observations in the data).

Another problem is related to the fact that the model is structured (nested). Such models may be estimated sequentially or simultaneously. It is desirable to estimate all levels simultaneously to avoid a bias in the calculated variance of the parameter estimates and to use data more efficiently. However, the number of alternatives may then become prohibitively high. There may also be other effects.

If the whole model is not estimated simultaneously, simultaneous estimation of various combinations of some of the choice levels may be thought of. However, frequency is related to a specific period, and the survey included trips for two different periods (trips > 100 km last 2 weeks and trips > 400 km last 6 months). As it is impossible to estimate one frequency model based on both types of frequency, the frequency model was restricted to include the frequency for trips > 100 km. For the other choices, all trip data were used (including, of course, destination choice sets corresponding to the trip category). Concerning the estimation of the rest of the structure, either the access choice or the egress choice will have to be separately estimated because they are assumed to be independent choices. Here, both choices were separately estimated, and, due to time limits, no tests were conducted to determine the effects of incorporating either of them into the mode and destination choice part of the structure.

Thus, access and egress models were estimated separately, the mode and destination choice simultaneously, and the frequency model separately. All levels are still connected by logsum variables. The simultaneous estimation also requires software that can accommodate such a complication (ALOGIT was used in this project).

A third problem is related to the fact that destination alternatives need to be described in terms of size. In this case, multiple sizes variables were used in the context of private trips, requiring specific capability of the estimation software.

DATA

Travel Survey

The data source is a national travel study conducted in 1984–1985. The interviews were individual home interviews spread out over the whole year. The total sample amounted to 7,600 persons. The rate of nonresponse was approximately 15 percent, yielding 6,500 individuals to be analyzed. The survey included long-distance trips as well as short-distance trips. Initially, the destinations for long-distance trips were not coded at a detailed level. A more detailed coding was introduced after the survey had been in process on for some time (for approximately 3,000 observations). These observations were used in the analysis.

The information that was collected included socioeconomic data for the individual and his or her household as well as trip-related information, such as access and egress modes, main mode, destination (at the 2,200 zone level), trip purpose, party size, number of overnight stays, and type of accommodations.

Transportation System Data

For each destination alternative (the chosen destination and sampled destination alternatives), data on travel time components were provided by the National Transportation Council, using a network analysis system (EMME/2). The car, train, air, and long-distance bus networks were coded at a level of detail corresponding to a subdivision of 504 zones. The difference between this zonal subdivision and the one used to define destination alternatives concerned mainly small agglomerations. Data for access and egress were taken from a special data base containing regional and local level-of-service data at the 2,200-zone level.

The construction of the mode-related cost variables had to rely on assumptions regarding the time of day of the trip as well as the mix of people in the traveling party, because this information was not included in the travel study and the discount systems for train as well as air were based on these factors. Also, overnight costs had to be calculated in many cases.

Data Describing Destinations

For each destination, data on the number of employees in different sectors of the economy were available. Also information on the population and area was available. For business trips, the number of employees in a subset of sectors was used. For private trips, the total population, the number of employees in the recreational sector, and the population density were used. Also, data on population density were used.

MODELS AND BUSINESS TRIPS

Access and Egress Mode Choice

For the choice of access modes to the station or the airport, four modes were defined. For egress, the number of modes is the same, but they are defined slightly differently. The modes for access and egress are nonmotorized modes, car, public transport, and taxi. The car mode was defined differently for access and egress, the obvious difference being the possibility to use a household car at the origin.

Separate models for access and egress trips were estimated. The parameters and the associated *t*-values of the model are presented in Table 1 for access as well as egress trips.

TABLE 1 Parameter Estimates and *t*-Values for Access and Egress Mode Choice Models—Business Trips

	Access		Egress	
Variable	Parameter	t-value	Parameter	t-value
Constant - walk	-0.7338	1.9	-0.03685	0.1
Constant - car	-2.031	2.0	-2.125	3.8
Constant - taxi	-2.159	3.6	-1.745	3.6
Car in household - car	3.028	2.9		
Household income - car			0.01037	3.9
Household income - taxi	0.01131	4.5	0.01108	5.0
Woman - taxi	1.304	3.8		
Cost	-0.002867	3.0	-0.003345	4.4
Time	-0.002026	2.5	-0.009894	2.5
Number of observations	300		283	
Log likelihood (parameters=0)	-401.79		-389.15	
Final log likelihood	-294.73		-293.46	
ρ^2	0.266		0.246	

Note: Income is in thousands of Swedish crowns per year before tax; cost is in Swedish crowns; time is in minutes per round trip.

The models exhibit approximately the same sensitivity to costs at the origin as at the destination. The sensitivity to time is, however, radically different, with a much greater sensitivity at the destination. A possible explanation is that the time spent at the origin does not have much alternative use as working time, because the access trip often takes place in the morning or evening, whereas the time at the destination often takes place during work hours.

The probability to use the more expensive modes is most likely related to the position of the traveler in the hierarchy and the economic strength of the company (or equivalent) where the person works. This is probably reflected in the salary of the person. However, person income was not reported in the survey, and household income is used as a proxy. Still, the effects are significant.

Choice of Main Mode and Destination

The parameter values for the mode and destination choice model are presented in Table 2. The model includes variables related to modes as well as to destinations. The model is simultaneously estimated, although with some important restrictions. Generally, simultaneous estimation is preferable to sequential estimation. In this case, simultaneous estimation increases the correlation between time variables, resulting in difficulties in estimating mode-specific time parameters.

Because a mode choice model could be estimated, the time and cost parameters were used as input to the estimation of the mode and destination model, scaled by a specific "scale" parameter. The parameter values from the mode choice model are reported with *t*-values in brackets because they are not estimated in the mode and destination model. The scale parameter, by which these made choice model parameter values should be multiplied, is reported separately with its associated *t*-value. The scale parameter is not significantly different from 1.

The cost parameters are segmented with regard to the type of worker. Full-time, salaried employees are likely to have higher values of time than others, which is reflected in the lower cost parameter for this category. The in-vehicle time parameter is much lower for train and bus as compared with car and air, which appears reasonable because working conditions are more favorable on trains and buses than in cars and aircraft. This was also found by Ridout and Miller (*4*). Waiting time has a significant influence if the frequency is higher than one train per 4 hr (in both directions).

The model also includes logsum parameters from the access mode model and from the egress mode model. The former is restricted to 1 because it otherwise would be larger than 1,

TABLE 2 Parameter Estimates and *t*-Values for Mode and Destination Choice Model—Business Trips

	Model 1		Model 2	
Variable	parameter	t-value	parameter	t-value
Constant - train	-2.898	4.4	-1.616	3.1
Constant - air	-3.807	5.1	-2.564	4.2
Constant - bus	-5.158	4.4	-6.024	5.3
In-vehicle/transfer time, car/air	-0.0024	(5.8)	-0.0024	(5.8)
train/bus	-0.0014	(4.9)	-0.0014	(4.9)
Cost, full time salaried employees	-0.00071	(3.2)	-0.00071	(3.2)
Cost, others	-0.0013	(5.3)	-0.0013	(5.3)
Wait time, train/air < 240 min	-0.0043	(2.5)	-0.0043	(2.5)
Parameter for generalised cost	1.090	10.3	1.083	10.4
Access (logsum)	1.0	-	-	
(distance, km)			-0.01183	2.8
Egress (logsum)	0.4912	3.5		
(distance, km)			-0.01421	2.7
Car in household - car	1.356	2.2	0.4306	0.8
Licenses per car - car	-0.5038	2.2	-0.5547	2.5
Travelling party > 4 persons - bus	3.152	2.5	3.269	2.6
Destination in Stockholm - air	0.8568	3.4	0.9564	3.5
Destination in smaller towns - air	-0.6861	2.2	-0.7420	2.4
Origin in Stockholm - air	1.165	4.5	1.395	5.2
Origin in medium sized towns - train	0.9884	3.9	0.9679	3.8
For all modes:				
Destination in Gothenburg	-0.08027	0.3	-0.07147	0.3
Destination in medium size towns	0.2640	1.4	0.2896	1.5
Destination in smaller towns	0.05974	0.3	0.09827	0.4
Destination in villages	-0.006948	0.0	-0.01682	0.1
Destination in rural areas	0.3109	0.9	0.3176	0.9
Size of destination (log of employees)	1.0	-	1.0	-
Logsum from mode choice	0.8410	8.0	0.8476	7.9
Number of observations	527		527	
Log likelihood (0)	-2267.48		- 2267.48	
Final log likelihood	-1472.33		-1483.52	
ρ^2	0.351		0.346	

although not significantly. These parameters make the choice of the main mode sensitive to changes in times and costs for access and egress modes. An alternative model, with the only difference that access and egress are represented by the distance, is also shown in Table 2. The alternative model has a ρ^2 of 0.352 compared with a ρ^2 of 0.346 for the base model, indicating that the probability that the alternative model is superior is low [in this case < 0.0001, using a modified likelihood ratio index test (7)]. The alternative model, however, has the advantage not to require information on access and egress modes, which can be unnecessarily demanding when access and egress modeling is not required.

The destination variables consist of a size variable and some dummy variables. The size variable parameter is constrained to 1. Thus, the probability to choose a destination is proportional to its size (other things being equal). The logsum parameter from the main-mode choice level to the destination choice level is significantly different from 0, but not from 1.

Choice of Frequency

The frequency model concerns the frequency of trips longer than 100 km (single distance). It includes a variable for the expected utility from such trips, measured as the logsum from the levels below (i.e., the destination, main mode, and access and egress levels). Zero frequency does not necessarily indicate nonmobility; it may well be the case that a number of shorter trips has taken place. Therefore, the model also includes a measure of the attractivity of such trips, namely the logsum of destination zones within 100 km. However, this logsum measure is based on a destination choice model, containing only a distance parameter and size variables.

Both of these logsum variables get significant parameters, which means that accessibility influences trip frequency. However, this does not necessarily prove a causality, because it may also be the case that workplaces of employees with high trip frequency locate where accessibility is high. The effect of, for instance, reduced travel costs on trip frequency may therefore be less than is predicted by the model.

The frequency model also includes the socioeconomic variables and dummy variables for type of origin zone. The estimated model parameters are presented in Table 3.

TABLE 3 Parameter Estimates and *t*-Values for Frequency Model—Business Trips > 100 km

Variable	Parameter	t-value
Constant - travel > 100 km	-6.069	3.4
Logsum > 100 km - travel > 100 km	0.6613	4.7
Logsum < 100 km - no travel > 100 km	0.4585	3.6
Woman - no travel > 100 km	1.116	4.4
Full time salaried employee - travel >100 km	0.9393	4.0
Age 24-45 - travel >100 km	0.5822	2.6
Origin Stockholm - travel > 100 km	0.7739	1.6
Origin Gothenburg - travel > 100 km	0.3828	0.7
Origin medium size towns - travel > 100 km	-0.4807	1.4
Origin in small towns - travel > 100 km	-0.7486	2.0
Origin in villages - travel > 100 km	-0.1116	0.4
Number of observations	1595	
Log likelihood(0)	-1105.56	
Final log likelihood	-329.36	
ρ^2	0.702	

MODELS FOR PRIVATE TRIPS

Access and Egress Mode Choice

For the choice of access modes to the station or the airport, the same four modes were defined as for business trips. Obviously, the possibility of being met at the station or airport by someone having a car depends on the trip purpose. Therefore, a dummy variable was introduced for the car alternative for the trip purpose "visit friends or relatives." Separate models for access and egress trips were estimated. The parameters and the associated *t*-values of these models for private trips are presented in Table 4.

For private trips, the access and egress models include some mode-specific dummy variables for origin and destination, respectively. These account to some extent for lack of information on distances, times, costs, and frequencies for the within destination zone part of the access and egress trips.

In both models, waiting time (half headway) and the time parameters differ significantly from 0, the magnitude of the parameters being slightly larger in the egress model. In both models, the waiting time parameter is less than the time parameter (which is equal for all modes). This is contrary to conventional wisdom concerning local trips, and may be because airport and train station services are often adjusted to departure times when frequencies are low.

The cost variable does not quite reach normal significance levels in the access model and is omitted in the egress model. The low-cost sensitivity may be due to other factors, such as time restrictions, the need to carry luggage, and, especially at the destination, lack of information on the local public transport system. It may, of course, also be due to the general coarseness of the model.

Mode and Destination Choice

As was the case for business trips, there were difficulties in estimating time parameters. Here it appeared obvious that attractive destinations (which are often small places) covaried with poor public transport service. Because the variables in the model can be expected to explain attractivity only to some extent, such a covariation can be expected to bias mode-related parameters. Therefore, these parameters were first estimated in a mode choice model and then included in the simultaneously estimated mode and destination choice model adjusted by a scale parameter. In this case, this parameter is also not significantly different from 1. The parameters for the time and cost variables indicate that in-vehicle time for the train is much less onerous than in-vehicle time for other modes, including railcar. The parameters of the model are shown in Table 5 (Model 1).

Also in this case there has been a segmentation of the cost parameter related to household income. The observations have been classified into two groups, with an income of 120,000 Swedish crowns (SEK) (1985 prices) as a divider. The cost

TABLE 4 Parameter Estimates and *t*-Values for Access and Egress Mode Choice Models—Private Trips

Variable	Access Parameter	Access t-value	Egress Parameter	Egress t-value
Constant - walk	-0.9478	3.6	-0.2758	1.0
Constant - car	-0.8457	3.1	-0.2802	1.1
Constant - taxi	-1.502	6.0	-1.386	4.9
Origin in Stockholm - public transport	1.052	3.3	-	
Origin in rural areas - public transport	-0.9976	2.0	-	
Destination in Stockholm - public transport	-		1.420	4.5
Destination in Stockholm - taxi	-		1.090	2.5
Destination in Gothenburg - public transport	-		1.102	2.8
Trip purpose to visit friends/relatives - car	-		1.126	4.6
Car in household - car	1.468	5.6	-	
Waiting time - public transport	-0.001051	3.1	-0.001454	2.8
Cost	-0.003371	1.8	-	
Time	-0.002412	3.0	-0.002924	3.0
Number of observations	385		342	
Log likelihood(0)	-525.95		-470.08	
Final log likelihood	-357.73		-344.86	
ρ^2	0.312		0.266	

sensitivity of the high-income group is only half the sensitivity of the low-income group.

The access and egress logsum variable is also included in the model. As for the business models, an alternative model using access and egress distance has been tested (Model 2 in Table 5). The differences between the models are small, also in terms of log likelihood. The model with the logsum variable is therefore not superior in terms of goodness of fit, but it provides the opportunity to calculate the effects of changes in times and costs of access and egress modes on main mode choice.

The destination variables include one multiple-size variable (total population and number of employees in the recreation sector) and a population density variable. Clearly, these variables cannot fully differentiate between different destinations for the mix of private-trip purposes. Some additional dummy variables indicate that trip purpose and time of year play a role for destination choice as well as mode choice.

The logsum parameter from main mode choice to destination choice also is not significantly different from 1 in this case.

Frequency Choice

The frequency model for private trips is similar to the one for business trips. As for business trips, the accessibility variables for trips outside and inside the 100-km border get significant parameter estimates (Table 6), although these estimates are lower than those for business trips.

The model also includes socioeconomic variables at the individual as well as the household level. At the individual level, the model includes the age of the interviewed person. The traveling party may, of course, include persons of different ages as well. At the household level, household income, summer house ownership, and the number of children are included.

VALUES OF TIME

Values of time are implicit in the models and take the form of estimated cost- and time-parameter values. For business trips, the values range from 40 SEK (approximately $6 U.S.) per hour (access trip) to 200 SEK (approximately $30 U.S.) for car and air trips for full-time, salaried employees (1985 prices). For private trips, there is a similar range, although the mean values are lower than those for business trips. For example, the value of in-vehicle time for private trips by train is about 60 percent of the value for business trips (60 SEK and 100 SEK, respectively).

The values of time implicit in the reported models are much higher than similar values found in urban studies, which normally range from 15 to 25 SEK for in-vehicle time. For the train, this is supported to some extent by stated preference studies, but it should be kept in mind that the cost variables are associated with considerable uncertainty. Therefore, the values of time should not be used in economic project evaluations until confirmed by other studies.

MODE CHOICE MODEL SPECIFICATION TESTS

Sweden is approximately 2,000 km from the south to the north, thus allowing a wide range of possible travel distances. Because longer distances will be associated with extra overnight stays for ground modes, this is a source of specific modeling difficulties. As described earlier, this has been, to a certain extent, accounted for in the model, but it can still be argued that the variance in the stochastic component in the utility functions is larger for longer trips (other factors may also contribute to this, such as more binding time constraints for ground modes on longer trips). This would violate the assumptions of the multinominal logit model, which requires the variance to be constant for all alternatives.

Therefore, a test was conducted to investigate whether there are significant differences in the variance for mode choice alternatives according to trip length. One way to test such a phenomenon would be to estimate relative scale factors for the utility functions for alternatives belonging to different trip-length categories and determine if they differ significantly. This is equivalent to estimating separate models for different categories, with the restriction that the parameters be the same up to a single-scaling factor. If this factor is less than 1 for a specific (distance) category, it suggests that the variance for the stochastic part of the utility function is larger for this group, because the scale parameter of the logit model is inversely proportional to the square root of the variance [see, for example, work by Ben-Akiva and Lerman (*1*)]. Such a test can easily be conducted using software that can simultaneously estimate a tree logit model (ALOGIT was used in this case).

The test was conducted as follows. The sample for the mode choice models used as input in the joint mode and destination choice model described above was subdivided into four groups according to distance. Then the same specification of this model was estimated using the full sample, but allowing for a separate scale factor (affecting all parameters in the utility function) for each of the subgroups except one (the reference group). The category for trips from 100 to 300 km was used as reference group. Each of the other groups thus had a specific scale parameter that could be tested statistically to see whether it differed from 1. The scale parameters and their

TABLE 5 Parameter Estimates and *t*-Values for Mode and Destination Choice Model—Private Trips

Variable	Model 1 parameter	Model 1 t-value	Model 2 parameter	Model 2 t-value
Constant - train	-1.044	3.2	0.1915	1.2
Constant - air	-2.434	9.0	-0.9014	3.6
Constant - bus	-1.011	4.9	-1.085	5.3
Invehicle time car/bus/jetplane	-0.002676	(12.3)	-0.002676	(12.3)
Invehicle time train (normal/sleep)	-0.001117	(4.3)	-0.001117	(4.3)
Invehicle time train - railcar	-0.002727	(3.6)	-0.002727	(3.6)
Invehicle time air (prop. aircraft)	-0.003415	(1.8)	-0.003415	(1.8)
Cost, household income <120 000 SEK	-0.001762	(7.3)	-0.001762	(7.3)
Cost, household income >120 000 SEK	-0.0008363	(2.9)	-0.0008363	(2.9)
Waiting time (half headway)	-0.002197	(2.0)	-0.002197	(2.0)
Number of transfers - train	-0.2512	(4.5)	-0.2512	(4.5)
Number of transfers - air	-0.3729	(2.9)	-0.3729	(2.9)
Scale parameter	1.061	18.6	1.106	19.3
Access/egress logsum	0.5324	4.7		
Access/egress distance			-0.01162	4.5
Car in household - car	2.088	12.5	1.913	11.8
Licenses per car - car	-1.804	8.9	-1.724	8.5
Age < 18 years - car	-0.9010	5.8	-0.8962	5.7
Age > 64 years - air	-1.295	3.1	1.393	3.4
Trip purpose recreation - car	0.4961	3.5	0.6939	5.1
Trip purpose recreation - bus	1.780	8.6	1.934	9.4
Trip purpose summer house - car	0.9862	3.3	1.147	3.8
Destination in Stockholm	-0.5552	3.2	-0.4169	2.5
Destination in Gothenburg	-0.1770	0.9	-0.1191	0.6
Destination in medium size towns	0.2063	1.3	0.1944	1.2
Destination in smaller towns	0.1201	0.9	0.1135	0.8
Destination in villages	0.1925	1.7	0.1927	1.7
Recr. trip in June/July - villages	0.4852	3.3	0.4889	3.3
Visit trip in June/July - rural areas	-0.5104	2.0	-0.5152	2.0
Population density in dest. zone	-0.04371	6.8	-0.04390	6.8
Logsum from mode choice	0.8912	16.3	0.8620	16.6
Size variables:				
Population	1.	-	1.	-
Number of employees in recreation branch, for recreation (not exponentiated)	4.837	25.1	4.841	25.1
Number of observations	1846		1846	
Log likelihood (0)	-7492.01		-7492.01	
Final log likelihood	-5733.41		-5733.23	
ρ^2	0.234		0.234	

associated standard errors are shown for business trips as well as for private trips in Table 7.

The meaning of these scale parameters is that the estimated parameter values (not shown here) are to be multiplied by these factors when applying the model to mode choice alternatives in a certain distance category. This means that the sensitivity to variable changes will be larger when the scale parameter is larger than 1, and reverse (everything else being equal).

As shown in Table 7, there are large differences between the different subgroups in the business model, although only the scale parameter for the third category is significantly different from 1. For private trips, the differences are not large, and none of the scale factors is significantly different from 1.

The results suggest that it is reasonable to include the full range of travel distances in the mode choice model for private trips (with the current specification), and that the model for business trips needs to be improved to meet the requirements for the multinomial logit model. These results may have an interest per se, although the specification of the joint mode and destination choice model (or the other models) was not analyzed in this particular way.

Further complexity of the model structure was also not tested within the reported project. The data are, however, still subject to research. A specification test that was conducted later (suggested by a referee) split mode choice into two levels: (a) the choice between the car and shared modes and (b) choice between shared modes. Although significantly

TABLE 6 Parameter Estimates and *t*-Values for Frequency Model—Private Trips > 100 km

Variable	Parameter	t-value
Constant - 1+ trips	-2.057	2.0
Logsum trip > 100 km - 1+ trips	0.1566	2.2
Logsum trip < 100 km - no trip	0.2453	4.4
Household income - 1+ trip	0.002457	2.5
Origin in Stockholm - 1+ trip	0.4212	1.9
Origin in rural areas - 1+ trip	-0.4648	3.0
Age < 19 years - 1+ trip	0.3538	1.8
Age 19-24 years - 1+ trip	0.5264	3.0
Age > 64 years - 1+ trip	-0.3635	2.1
Number of persons < 12 years in household - 1+ trip	-0.1034	1.4
Household owns summerhouse - 1+ trip	0.5725	5.1
Number of observations	2700	
Log likelihood(0)	-3196.57	
Final log likelihood	-1182.73	
ρ^2		0.630

TABLE 7 Scale Parameter Estimates and Standard Errors for Distance Groups

Travel distance	Business trips		Private trips	
(single way)	scale parameter	std error	scale parameter	std error
Up to 300 km	1.0	-	1.0	-
301 - 600 km	1.382	0.258	0.8897	0.0706
601 - 900 km	0.6250	0.140	0.8985	0.0793
901 -	0.8692	0.214	0.8844	0.102

better in terms of the likelihood ratio test, such a structure implied poor cost parameter estimates for business trips and affected the parameter estimates of the private trips model only marginally (the logsum parameter being 0.8).

CONCLUSIONS

Long-distance travel behavior is treated as individual choices of trip frequency, destination, main mode, and access and egress modes. A system of structured logit models was estimated for these choices. Separate models were estimated for business trips and private trips. The model exercise shows that long-distance travel behavior is sensitive to the following:

- Socioeconomic characteristics of the individual and of the household,
- Characteristics of the destination in terms of population and employment,
- Characteristics of main modes, and
- Characteristics of access and egress modes.

The model exercise further shows that these characteristics are influential at all choice levels. The relative importance of these characteristics is reflect in the model parameters. Specifically, train in-vehicle time seems to be less onerous than in-vehicle time for other modes. Also, cost sensitivity seems to be quite different among types of employes and among household income groups.

Long-distance travel behavior is, of course, more complicated than is reflected in the model system. Among the neglected behavioral phenomenon are trip chaining and the use of different modes on outbound and homebound trip legs. Also, the models were estimated using a travel study that was not specifically designed for such a task, yielding less accurate information than would have been desirable and making it impossible to account for time availability.

However, modeling long-distance travel behavior by using discrete choice models seems to be a viable way to achieve a tool for evaluating infrastructure investment and other changes of the transportation system.

ACKNOWLEDGMENT

This analysis was funded by the Swedish Transportation Research Board.

REFERENCES

1. Ben-Akiva, M., and S. Lerman. *Discrete Choice Analysis.* MIT Press, Cambridge, Mass., 1985.
2. Stopher, P., and J. Prashker. Intercity Passenger Forecasting: The Use of Current Travel Forecasting Procedures. *Transportation Research Forum, Proc., 17th Annual Meeting*, 1976, pp. 67–75.
3. Koppelman, F. Models of Intercity Travel Choice Behavior. Presented at 68th Annual Meeting of the Transportation Research Board, Washington, D.C., 1989.
4. Ridout, R., and E. Miller. A Dissaggregate Logit of Model of Intercity Common Carrier Passenger Modal Choice. *Canadian Journal of Civil Engineering*, Vol. 16, No. 4, 1989, pp. 568–575.
5. Wilson, F., S. Damodaran, and D. Innes. Disaggregate Mode Choice Models for Intercity Passenger Travel in Canada. *Canadian Journal of Civil Engineering*, Vol. 17, No. 2, 1990, pp. 184–191.
6. Ramjerdi, F., and A. Madslien. Long Distance Travel Choice Behavior, the Case of Norway. Presented at the 6th International Conference on Travel Behavior, Quebec, 1991.
7. Horowitz, J. L. Statistical Comparison of Non-Nested Probabilistic Discrete Choice Models. *Transportation Science*, Vol. 17, No. 3, 1983, pp. 319–350.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.