# LONG DISTANCE TRAVEL BEHAVIOUR MODELS IN SWEDEN

**Staffan Algers**
**Royal Institute of Technology/ TRANSEK AB**
**Stockholm - Sweden**

## INTRODUCTION

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 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 big need to be able to analyse how changes in prices and level of service influence ridership, also with regard to expected changes in the economic activities over a forecasting period.

In the last years, forecasting of such changes have been based on a linked model system comprehending 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 utilise the advantages of disaggregate modelling by using it for all steps in an integrated structure. Such a step has now been taken, and the scoop of this paper is to give an overview of the models involved in the new model system. The analysis has been carried out by TRANSEK AB, commissioned by the Swedish Council of Transportation and funded by the Swedish Transportation Research Board.

The concept of "long-distance travel" is not very well defined, although it is frequently used. Obviously, the notion is used to define a specific category of trips. Various criteria for this can be thought of, of which the trip length is one - but not necessarily the most adequate one for modelling. However, given such a criterion in the 1984/ 85 National Travel Survey in Sweden, there was little choice left for the modelling work. In this survey, trips longer than 100 km (one direction) were identified as long-distance trips.

## 1. ANALYSIS OF LONG DISTANCE TRAVEL BEHAVIOUR - AN ANALYSIS OF DISCRETE CHOICE

As indicated above, the purpose of the modelling effort was to produce a system of forecasting models, comprehending mode split, trip (spatial) distribution and trip generation. In the previous analysis, mode split was analysed using probabilistic discrete choice models, specifically the well known logit model (Ben-Akiva and Lerman 1985). This approach was adhered to also when extending the model system to trip distribution and trip generation. Discrete choice analysis has been applied to long distance travel also in other studies (Koppelman, 1989, and Ramjerdi, 1991), but, to the knowledge of the author, not comprising access/egress mode choice, main mode choice, destination choice and frequency choice in an integrated structure.

The demand for long distance trips is viewed as the result of the behaviour of utility-maximising 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 travelling 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 go back to their homes again. 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.

### 1.1 Alternatives

The mode alternatives for long distance trips were defined to consist of four main modes - 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 very rare. The utility of the train, air and bus modes may depend on the possibilities to get to/from the train/bus station and to the airport, at the origin as well as at the destination. Since the access/egress modes may be of interest as policy variables, and since data permitted, the access/egress alternatives were also modelled as separate alternatives.

The destination alternatives were defined to be approximately 2200 agglomerations and rural areas, adding up to 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 analysed period or not. The fraction having made more than one trip was very small.

### 1.2 Variables

Most variables that are included in the analysis can be grouped into three main classes - time and cost variables relating to the access/egress and main modes, size variables relating to destinations and socio-economic variables relating to the travellers.

## 2. STRUCTURE OF A LOGIT MODEL FOR LONG DISTANCE TRAVEL BEHAVIOUR

It is assumed that the alternative that is associated with the largest utility is chosen. This alternative may be not to travel as well as a combination of mode and destination. Analytically, the utility is defined to consist of one part that is observed by the analyst, and one unobserved part. The assumption of a stochastic distribution of the unobserved part makes it possible to model the probability of a certain alternative having the largest utility to the choice maker.

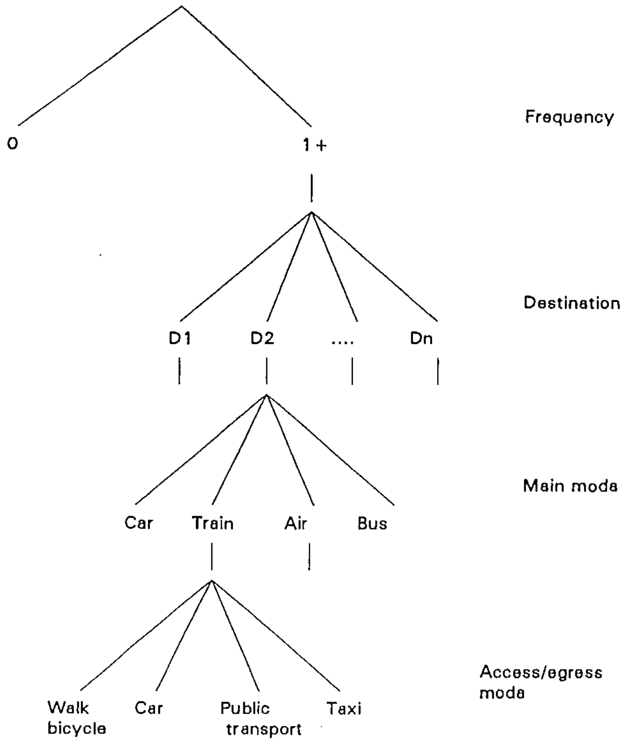
In the logit model, it is assumed that the stochastic part of the utility function is identically and independently distributed. This assumption may not be sufficiently realistic, especially when the model includes alternatives of different kinds, such as mode, destination and frequency alternatives. In order to avoid (or at least reduce) this problem, the alternatives can be structured in different levels, within each of which the assumption is assumed to hold. The model is then called a structured (or nested) logit model. The structure is empirically derived, and depends on the specification of the model.

In figure 1, the general structure of the model is shown. The choice of access and egress modes is positioned at the bottom of the model. The actual structure is somewhat simplified in the figure, since the choice of access and egress modes is treated as two independent choices. At the next level is the choice of the main mode. At this stage, the choice of the main mode is influenced by the accessibility to the airport/station given by the logsum variable from the access/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/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 lower levels will also affect all the higher levels.

As an example, we can think of an improvement of a bus service to an airport. This 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, since 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 (since the airport is easier accessible), which will cause a shift to these destinations from other destinations. Finally, since accessibility is generally improved trip frequency will also increase. The improvement of the bus service will thus influence all choices in the structure.

Graph 1 General model 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 that is under consideration. This sensitivity is embedded in the parameters of the model, which have been estimated by the use of statistical software.

## 2.1 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.

### 2.2 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 very 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 can be estimated sequentially or simultaneously. Principally, it is desirable to estimate all levels simultaneously, to avoid a bias in the calculated variance of the parameter estimates, and to utilise data more efficiently. However, the number of alternatives may then become prohibitively high. There may also be other effects, which are discussed below. That lead to the decision to estimate the access and egress models separately, the mode and destination choice simultaneously and the frequency model separately. All levels are still connected by logsum variables. The simultaneous estimation requires software that is capable of such a complication (ALOGIT was used in this project).

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

## 3. DATA

### 3.1 Travel survey

The data source is a national travel study conducted in 1984/85. The interviews were individual home interviews, spread out over the whole year. The total sample amounted to 7 600 persons. The rate of non-response was approximately 15 percent, yielding 6 500 individuals to be analysed. A subset of these observations, which were coded at a detailed level was used in the analysis.

The information that was collected included socio-economic data for the individual and 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 accommodation.

### 3.2 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 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 for defining destination alternatives concerned mainly very small agglomerations.

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

### 3.3 Data describing destinations

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

## 4. MODELS FOR BUSINESS TRIPS

### 4.1 Access/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 slightly differently defined. The modes for access and egress are non-motorised modes, car, public transport and taxi. The car mode was differently defined 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 shown in table A1 in the appendix, for access as well as egress trips.

The models exhibit approximately the same sensitivity to costs at the origin as well 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, since the access trip often takes place in the morning or in the evening, whereas the time at the destination often takes place during working hours.

The probability to use the more expensive modes is probably related to the position of the traveller 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, personal income was not reported in the survey, and household income is used as a proxy. Still, the effects are significant.

### 4.2 Choice of main mode and destination

The parameter values for the mode and destination choice model are listed in table A2 in the appendix. 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 correlation between time variables, resulting in difficulties in estimating mode specific time parameters.

Since a mode choice model could well 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, since they are not estimated in the mode and destination model. The scale parameter, by which these parameter values should be multiplied, is reported separately with its associated t-value. The scale parameter is not significantly different from one.

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/bus as compared to car/air, which seems reasonable since working conditions are more favourable on trains/buses than in cars/aircraft. Waiting time has a significant influence if the frequency is higher than one train per four hours (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 one, since it otherwise would be larger than one, 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, where access and egress are represented by the distance is also shown in table A2. That model is clearly inferior in terms of log likelihood, but has the advantage not to require information on access / egress modes, which can be unnecessarily demanding when access/egress modelling is not required.

The destination variables consist of a size variable and some dummy variables. The size variable parameter is constrained to one. 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 zero, but not from one.

### 4.3 Choice of frequency

The frequency model concerns the frequency of trips longer than 100 km (single distance). It includes as a variable the expected utility from such trips, measured as the logsum from the levels below - the destination, main mode and access/egress levels. Zero frequency does not necessarily indicate non mobility - 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 only based on distance parameter.

Both of these variables get significant parameters, which means that accessibility has an influence on trip frequency. However, this does not necessarily proof a causality, since 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 socio-economic variables and dummy variables for type of origin zone. The estimated model parameters are listed in table A3 in the appendix.

## 5 MODELS FOR PRIVATE TRIPS

### 5.1 Access/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/airport of someone having a car is depending on the trip purpose - therefore, a dummy variable was introduced for the car alternative for the trip purpose "visit friends/relatives". Separate models for access and egress trips were estimated. The parameters and the associated t-values of the models for private trips are not shown here, because of the limited space.

For private trips, the access/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/egress trip.

In both models, waiting time (half headway) and the time parameters differ significantly from zero, 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 due to the fact that 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, need to carry luggage and - especially at the destination - lack of information of the local public transport system. It may, of course, also be due to the general coarseness of the model.

### 5.2 Mode and destination choice

As was the case for business trips, there were difficulties in estimating time parameters. Here, it seemed obvious that attractive destinations (which are often small places) covaried with poor public transport service. Since the variables in the model only to some extent can be expected to explain attractivity, such a covariation can be expected to give mode related parameters a bias. 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. This parameter is also in this case not significantly different from one. The parameters for the time and cost variables indicate that in-vehicle time for train is much less onerous than in-vehicle time in the other modes, including railcar.

Also in this case has there been a segmentation of the cost parameter, here related to household income. The observations have been classified into two groups, using an income of approximately 250 000 Swedish crowns in 1992 prices as a divider. The cost sensitivity of the high income group is only half the sensitivity of the low income group.

The access/egress logsum variable is also included in the model. As for the business models, an alternative model using access/egress distance has been tested. The differences between the models are very small, also in terms of log likelihood. The model with the logsum variable is therefore not superior in terms of goodness of fit, but gives the opportunity to calculate 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 can not fully differentiate between different destinations for the mix of private trip purposes. Some additional dummy variables indicate that trip purpose and time of year plays a role for destination choice as well as for mode choice.

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

### 5.3 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, although lower than for business trips.

The model also includes socio-economic variables at the individual as well as at the household level. At the individual level, the model includes the age of the interviewed person. The travelling party may, of course, include persons of different age as well. At the household level, household income, summerhouse ownership and the number of children are included.

## 6. CONCLUSIONS

Long distance travel behaviour is treated as choices of trip frequency, destination, main mode and access/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 showed, that long distance travel behaviour is based on

- socio-economic characteristics of the individual and of the household
- characteristics of the destination in terms of population and employment
- characteristics of main modes
- characteristics of access and egress modes

The model exercise shows further, that these characteristics are influential at all choice levels. The relative importance of these characteristics are reflected 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 between types of employees and between household income groups.

Long distance travel behaviour is, of course, much more complicated than is reflected in the model system. Among the neglected behavioural 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 take time availability into account.

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

## BIBLIOGRAPHY

Ben-Akiva, Moshe and Lerman, S.. Discrete Choice Analysis. MIT Press, 1985.

Koppelman, Frank. Models of Intercity Travel Choice Behaviour. Paper presented at the 68th Annual Meeting, Transportation Research Board. Washington 1989.

Ramjerdi, F and Madslien, A.. Long Distance Travel Choice Behaviour, the Case of Norway. Paper presented at the 6th International Conference on Travel Behaviour. Quebec 1991.

## APPENDIX: MODEL PARAMETERS

**Table A1 Parameters and t-values for access and egress mode choice models**

| VARIABLE | ACCESS | | EGRESS | |
|---|---|---|---|---|
| | 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 | |
| Likelihood(parameters=0) | -401.79 | | -389.15 | |
| Final likelihood | -294.73 | | -293.46 | |

Income is in thousands of Swedish crowns per year before tax, costs are in Swedish crowns and times are in minutes (round trip).

**Table A2 Parameter estimates and t-values for mode and destination choice model - business trips.**

| VARIABLE | Model 1 | | Model 2 | |
|---|---|---|---|---|
| | 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 |
| Licences 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 | |
| Likelihood (0)* | -2267.48 | | - 2267.48 | |
| Final likelihood | -1472.33 | | -1483.52 | |

*) Parameter for sampling correction = 1

**Table A3 Model parameters and t-values for frequency model - business trips**

| VARIABLE | Parameter | t-value |
|---|---|---|
| Constant - travel > 10 mil | -6.069 | 3.4 |
| | | |
| Logsum > 10 mil - travel > 10 mil | 0.6613 | 4.7 |
| Attraction < 10 mil - no travel > 10 mil | 0.4585 | 3.6 |
| | | |
| Woman - no travel > 10 mil | 1.116 | 4.4 |
| Full time salaried employee - travel >10 mil | 0.9393 | 4.0 |
| Age 24-45 - travel >10 mil | 0.5822 | 2.6 |
| | | |
| Origin Stockholm - travel > 10 mil | 0.7739 | 1.6 |
| Origin Gothenburg - travel > 10 mil | 0.3828 | 0.7 |
| Origin medium size towns - travel > 10 mil | -0.4807 | 1.4 |
| Origin in small towns - travel > 10 mil | -0.7486 | 2.0 |
| Origin in villages - travel > 10 mil | -0.1116 | 0.4 |
| Number of observations | 1595 | |
| Likelihood(0) | -1105.56 | |
| Final likelihood | -329.36 | |