Anne Madslien
am@toi.no

**Institute of Transport Economics**

PO Box 6110 Etterstad, N-0602 Oslo, Norway
Telephone No.: +47 22 57 38 00 Telefax No.: +47 22 57 02 90
http://www.toi.no

# Development of a Regional (short distance) Transport Model System in Norway

*Paper to be presented at Young Researchers Seminar, ECTRI-FERSI-FEHRL, The Hague May 2005*

# Introduction

Every fourth year the Transport Authorities in Norway present a National Transport Plan. The ambition is to include all types of transport in this plan, and the last plan focused on the period 2006-2015. Traditionally, infrastructure projects in different transport sectors are evaluated using different transport models, different methods for measuring the impacts of the projects, and different sets of base data. To be able to compare projects in the rail, road, air and sea sectors, the Ministry of Transport and Communications decided to develop a common tool for all transport sectors to be used for evaluation of projects and policies.

Based on this the Institute of Transport Economics (TOI), together with Molde Research Institute, were asked to develop a model system for travel demand in Norway. The clients of the work have been the Norwegian National Public Road Administration, the Norwegian National Rail Administration, AVINOR (AVIation NORway) and the Norwegian Coastal Administration. The Ministry of Transport and Communications has provided supplementary funding. A national long distance model (for trips longer than 100 km one way) has earlier been developed in Norway, and the last version, with 1400 zones nationwide, was completed and implemented in 2002. The newly developed regional model system covers trips shorter than 100 km one way. In the next National Transport Plan it will be mandatory for all the transport authorities to use the model, with common base data and assumptions. This will give a far better foundation for assigning priorities between projects in the different transport sectors in future Transport Plans than in the previous ones.

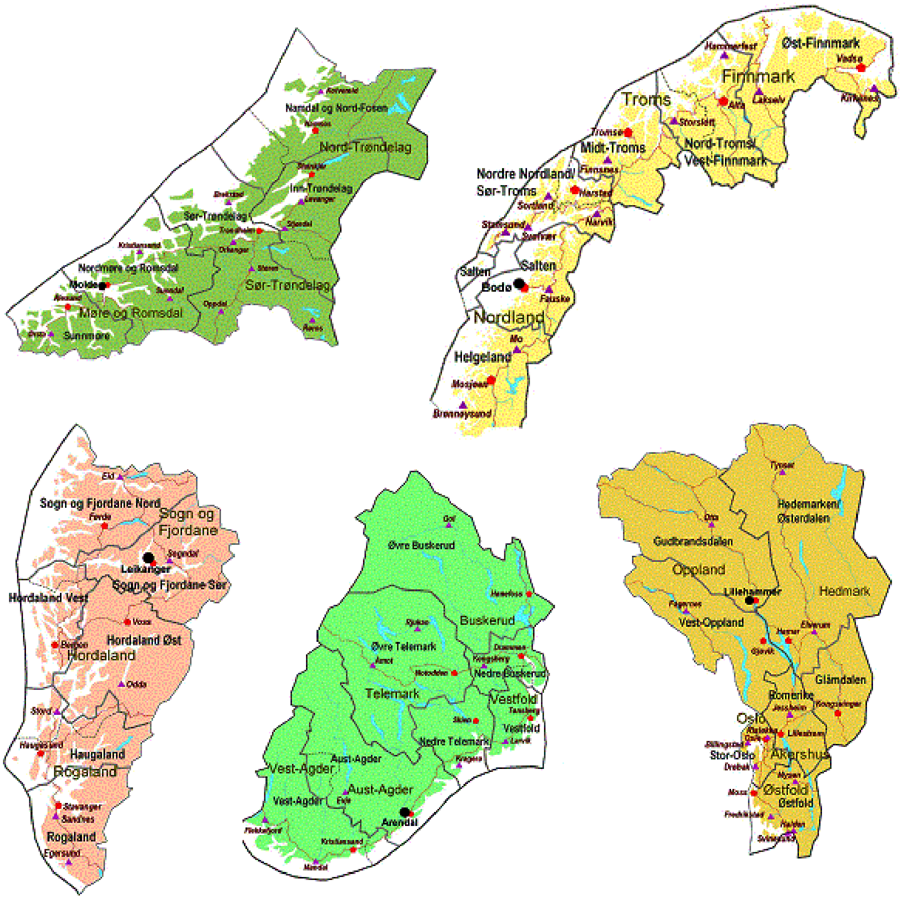
This paper gives an introduction to the basic elements in the new regional model system in Norway. Some of the paper is based on a paper written in Norwegian by Jens Rekdal (Molde Research Intitute) and Anne Madslien (TOI) for the conference *Trafikdage på Aalborg Universitet 2004*.

Molde Research Institute and TOI have been responsible for deciding the overall model structure, process the data, and estimate and implement all of the different sub models. Most of the estimation work is done by Odd I. Larsen and Jens Rekdal (Molde Research Institute) and Anne Madslien (TOI). The collection of data was the responsibility of our clients. The main aspects of the data collection was to conduct a comprehensive travel survey in Norway (NTS2001) and to establish the network models (which includes coded road network and public transport routes for the whole country) and producing data reflecting level of service of the networks. Some of this work was contracted out to consultants.

# The Model System

The model system consists of 6 transport models covering all of Norway. The country is divided into five regions (figure 1), and a sixth model is implemented for the Intercity-region (covering Oslo and its commuting area). Each model basically has the same structure, but they use different data sets and there are some geographical dummies for different types of regions.

*Figure 1. The five regions in Norway*

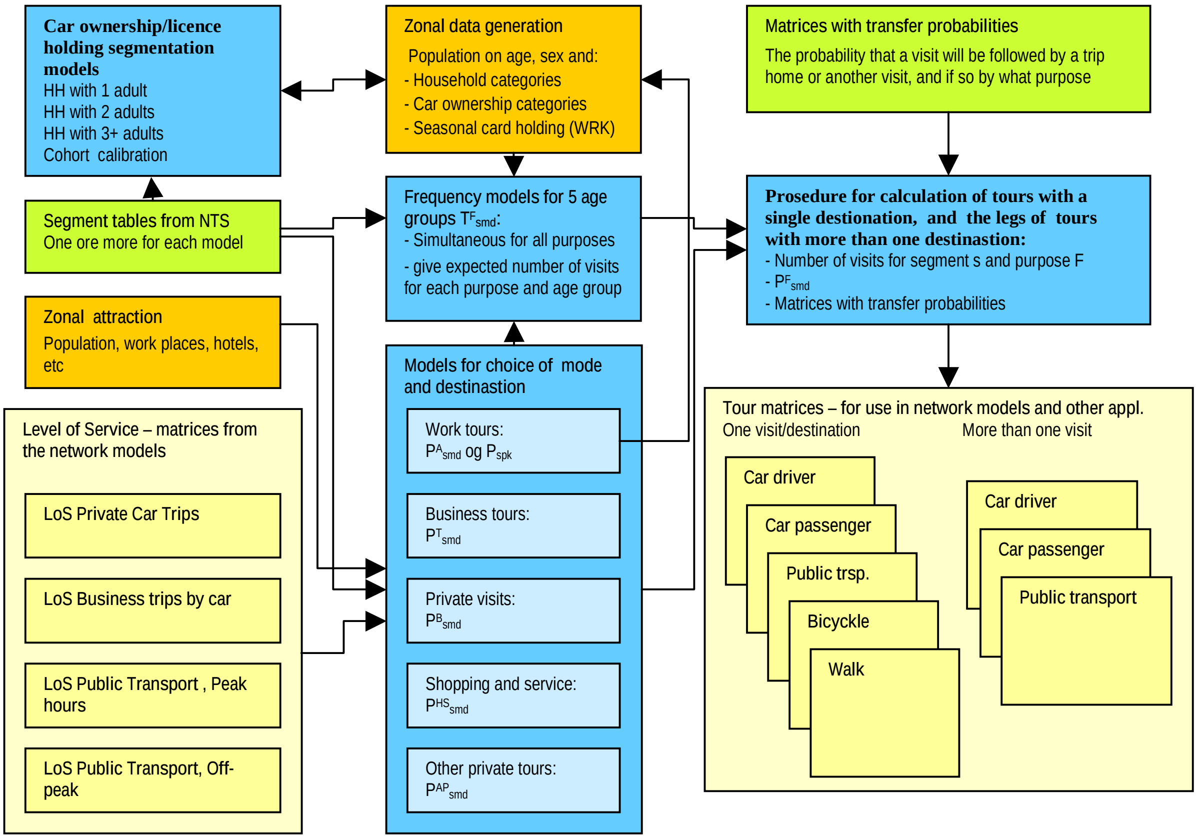



The general structure of the system of models is shown in figure 2.

A model run normally starts by producing *Level of Service (LoS) matrices* from the network models. These are matrices containing information about travel times and costs between all zones in the region and for all modes of transport.

The LoS-data form input to a run of the transport demand model, together with *demographic zonal data* for generation of trips, zonal data for attraction of trips and different *segment tables* from the National Travel Survey (NTS2001). These tables give average values for segments of the population for variables where zonal data do not exist in any public register. Data describing the *almost 14 000 zones nationwide* were mainly supplied by Statistics Norway.

*Figure 2. Structure of the model system*




Each of the models will be described in short in the following sections.

# Car ownership and driver license holding segmentation models

The demographic data available for the model system divides the population in the zones by age (5 years intervals), sex and three household categories:

- households with one adult (age 18 +)
- households with two adults
- households with three or more adults

The *car ownership and license holding situation* (together making *car availability* information used by other sub models) in the household are important aspects of an individual's possibility to make certain trips, and at the same time the changes in car availability over time probably are very important for the overall increase of car traffic. Therefore, we wanted our models to capture the effects of changes in car ownership and license holding. The purpose of these models is to segment the population further than age, sex, and household types, so that information about the car availability for different population groups can be used in all parts of the model system. Every group of a given sex, age interval, and household type, in a zone, is subdivided in 5 car availability segments:

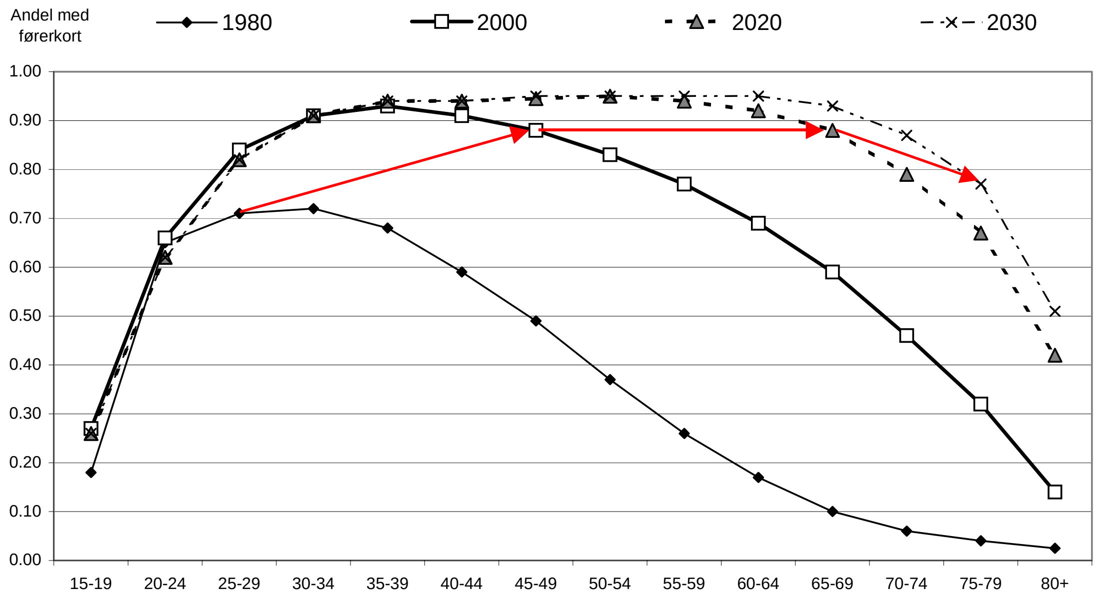
S1: Individuals without driver license and no car in the household.
S2: Individuals without driver license, but one or more cars in the household.
S3: Individuals with driver license, but no car in the household.
S4: Individuals with driver license and at least as many cars as licenses in the household.
S5: Individuals with driver license and less cars than licenses in the household.

People in the different segments will have different possibilities of making a trip by car, e.g. individuals in segment S4 have full car availability while those in segment S1 will have no accessibility to car as driver and bad accessibility to car as passenger.

For each of the three household categories, models dividing the population on the five car availability segments are estimated and implemented. The models are estimated using a maximum likelihood procedure coded in GAUSS. "Utility functions" are specified consisting of alternative specific constants and variables describing the zone of living and characteristics of the individual and the household (data from the National Travel Survey 2001). Important variables are age, sex, family type, population density and household income.

When implemented the models are calibrated to base forecasts (calculated by use of cohort effects) for license holding in the years 2010, 2015, 2020, 2025 and 2030. For women these time effects are illustrated in figure 3.

*Figure 3. Cohort effects in driver license holding. License holding rate for women.*




The red arrows between the curves indicate how the rate of license holding for women who were in the age group 25 to 29 in 1980, will change over time. In 2000 these women were between 45 and 49 of age, with a significant increase in license holding rate. As time moves on these women's drivers licence rate does not change much until 2020. Then it decreases, but the rate is still higher in 2030 than in 1980. When the models are used for long-term traffic forecasts the calibration will take care of these cohort effects.

# Models for choice of destination and mode

Five models are estimated to simulate the simultaneous *choice of mode and destination*, one model for each of the travel purposes:

- work trips
- business trips
- private visits
- shopping and service trips
- other private purposes

The main data source for estimation of the models has been the National Travel Survey 2001, where 20 000 respondents were interviewed. This data source was supplemented with data from a travel survey in the Oslo-region that was carried out in the same period and with a very similar questionnaire (8000 respondents). The major advantage of pooling the two surveys was a much better coverage of public transport trips. Both surveys include trip diaries for one day, together with information on background variables for the respondents. For both surveys a considerable effort was spent on accurate coding of the geographical origin and destination of trips. Some of the data were also recoded for use in the estimation of mode/destination models. A set of rules was designed and used to recode tour chains with more than one destination into tour/return trips from home to a single main destination and back. The rules were based on time spent on each destination, distance from home to each visited destination, total distance for the trip chain and total distance for any possible simplification of the trip chain. Complex chains with no distinct main destination/purpose and/or many destinations were excluded from the data set used for estimation. This left us with about 23 000 tours in the data set for estimation.

For each observation (round trip) a random sample of 249 alternative destinations (within 100 km from the residential zone) was drawn to allow for simultaneous estimation of mode and destination choice on a subset of alternatives. Zonal data for attraction and LoS-data were connected to both the chosen destination and each of the alternative (drawn) destinations.

The estimation work was a relative time-consuming process because of problems in the LoS-data, and new data had to be produced a couple of times. Unfortunately we know there are still some problems in the coding of both the road networks and the public transport lines. The clients of the project are working hard to increase the data quality before the final calibration of the model system.

The five mode/destination models are home- and tour based, and they use segmentation from the car availability and license holding models as direct input. The models are estimated for the following five possible travel modes/means:

- Car as driver
- Car as passenger
- Public transport (rail, subway, tram, boat and bus service)
- Bicycle
- Walk

Bicycle as mode of transport has a small market share nationwide in Norway, varying from 3 per cent for shopping and service trips to 8 per cent of the private visits. Information from the coded network for public transport indicates that a significant amount of the observations did not have public transport as an available mode for the actual trip (as much as 60 per cent for some of the purposes). We believe that this percentage is higher than actual, indicating that the network for public transport is not coded as detailed as desired.

The data show significant differences between men and women both when it comes to the choice of transport mode and the trip length distribution. Women are more often car passengers than men, and they more frequently use public transport. They also walk more than men seem to do. On average, women travel shorter distances than men, especially for work trips and business trips. One explanation to this is a complete different choice of occupation between men and women.

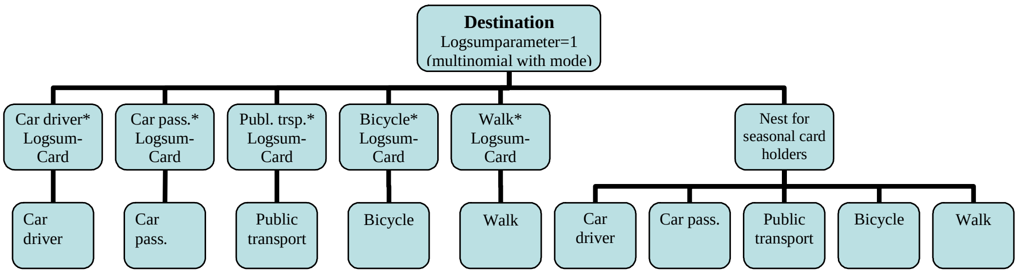
The models for mode and destination are logit models. The logit model describes how individuals choose between different alternatives (combinations of mode and destination), assuming utility maximizing. In a multinomial logit model the utility of each alternative is divided into a deterministic and a random component, and the random components of all alternatives are assumed independent and Gumbel distributed. We want to study the choice of mode and destination simultaneously, and in this case the assumption of independent random components may be unrealistic. We can then introduce structured (or nested) logit models, allowing for different random terms at different levels in the three-structure. In our work we have tested alternative tree-structures for all travel purposes. For most purposes the variant with mode choice on the upper level had a somewhat better log-likelihood-value than the multinomial model, but the value of the cost coefficient was significantly smaller and it had lower significance, which led to higher implicit values of travel time. Based on the fact that the values of time were already higher than expected, we chose to use the multinomial model for three of the purposes (business, private visits, shopping/service). The model for "other private trips" is structured with destination over mode, while the model for work trips has a special structure with the acquisition of seasonal card for public transport estimated simultaneously with the mode and destination choice.

The model for "other private trips" was the most difficult to estimate with satisfactory results, probably because it is a very heterogeneous travel purpose including leisure and recreational travel (often quite long trips), delivering and picking up other people (often very short trips, e.g. kids to and from kinder-garden) and several other small purposes.

The model for work trips has a somewhat more complicated structure than the other models, with simultaneous estimation of the choice of mode, destination and seasonal card (zero cost for marginal trips) for public transport. The simple (and usual) way of handle seasonal card in models, is to use the price of the card divided by number of days travelled as the cost for public transport for those having a seasonal card (and the same cost for all possible destinations). This has several disadvantages, most important is the fact that the cost of public transport often is perceived as zero at the registration day for seasonal card holders. According to the travel survey, a small portion of the seasonal card holders uses their car when they go to work at the day of registration. These travellers (usually going by public transport) rarely own discount cards for toll cost and ferries, making the perceived cost of choosing car even higher. In our work trip model these aspects are captured, in a far more correct way than the traditional approach in mode/destination models.

In the model for work trips seasonal card holding is placed in a separate nest with its own logsum parameter (figure 4). This implies that we get 5 additional utility functions, one for each mode. The other utility functions are multiplied by this same logsum parameter to bring them to the same level (a “dummy-nest” for each mode). This is necessary to get the correct generic parameters. The model is multinomial except for the nest for seasonal card (this is the reason for “logsumparameter=1” for destination in the figure).

*Figure 4. The model structure for work trips.*




In the nest reflecting alternatives having a seasonal card, the cost of using car as driver or passenger is defined as the sum of the ordinary cost of using the mode, and the per day cost of the season card. The cost of using public transport is the per day cost of the season card only. *Inside* the nest for seasonal card holders, public transport will be perceived as free, and will have a high probability of being chosen. *Between* the nests, public transport will be less expensive with a seasonal card than without. At the same time other modes will be more expensive for those having a seasonal card. The cost of seasonal card varies by origin and destination, and is most often dependent of the distance, but there are also unit prices for given regions (inside some of the cities) and zonal based prices (the larger Oslo-area).

There are several advantages using this approach modelling the choice of mode, destination and seasonal card simultaneously for work trips, e g:

- More correct specification of the choices, at the same time some of the correlation between travel time and travel cost is broken.
- The model for seasonal card holding can be used for the four other travel purposes as long as the model for work trips is run first.
- It becomes possible to analyse effects of different pricing schemes of single tickets and seasonal cards. Changes in prices will have an impact on both seasonal card holding and mode choice.
- It becomes possible to differentiate the costs for car drivers and car passengers depending on whether they use these modes sporadic or not (through discounts for toll costs and ferries). This will also break some correlations.

We will not show the complete models for mode/destination for the five travel purposes, but some “highlights” which are common for the models are worth mentioning:

- The coefficients associated with the waiting time variable for public transport were difficult to estimate, partly because of how this variable was calculated in the network models, and partly probably because of the problems with the network coding. Lots of different variants of the variable were tested, as “open” and “hidden” waiting time, waiting time per boarding, different transformations, the sum of waiting time and access time, etc. In the final models waiting time is transformed by the square root. The coefficients have the right sign and size, and is significant different from zero. This formulation implies that the wait time weight will vary by the time spent on waiting, and that the weight of long waiting times is reduced.
- A dummy introduced for observations with at least two destinations of the tour reflects increased probability of using car as compared to tours with one destination.
- The five different categories of car availability are important variables in all five models. In general full car availability implies increased probability for being a car driver. In households with more license holders than cars, women seem to loose the “battle of using the car”.
- Several dummies for age and type of region (large city etc) are tested and to some extent included in the models. However, we have tried to avoid having too many segments in the final models (many segments increases the computer time dramatically), and only kept the most significant ones.
- In Norway information about parking possibilities (number of places and associated costs) at a zonal level is very limited. We have therefore calculated indices based on the density of working places at the different destinations. These indices were meant to reflect how difficult (and maybe how expensive, or how far from the actual destination) it is to get a place to park the car in different zones. The different indices are introduced in the models as dummy variables. For most purposes (all except private visits) high density of working places reduces the probability of using car. The size and sign of the coefficients of these dummy variables, compared with the cost coefficients, reflect a reasonable perceived cost when the duration of the visits (parking time) for different travel purposes is taken into account.
- When estimating the models, there were some problems getting the models to reproduce the observed trip length distribution. The models underestimate the shortest and longest trips and overestimate the trips of medium length. Distance dummies had to be introduced to overcome this problem.

One interesting aspect of this type of models is to study and compare the size and sign of the most important variables in the models, and especially the ratio between the time coefficients and the cost coefficient (this reflects the implicit value of travel time savings). The implicit values of travel time savings seem a little high in all models, except for work tours. The main reason may be that the travel times used in estimation are calculated in a network model without capacity problems (giving too short travel times). There is also correlation between time and cost, which might give a tendency for one of the coefficients to reflect the total effect of distance. The values are particularly high for the category “other private trips”, probably because this travel purpose is very heterogeneous, from long recreation trips to short delivering/picking up trips. However, the majority of these trips takes place in weekends, when the value of time is considerable lower than for the rest of the week (see “in addition, weekend” in table 1). This can be a consequence of a “looser time budget” in the weekends. This tendency we also find for private visits. The implicit values of travel time from the models are shown in table 1.

*Table 1. Implicit values of travel time from the mode/destination models. NOK/hour (1 Euro = approx. 8.2 NOK)*

| | Work | Business | Private visits | Shopping/serv. | Other |
|---|---|---|---|---|---|
| **Car driver (CD):** | | | | | |
| Men | | 135 | 63 | 66 | 198 |
| Women | | 242 | 89 | 87 | 253 |
| In addition, large city | | 58 | | | |
| In addition, weekend | | | -19 | | -84(m), -77(f) |
| Men, less than 50 year | 55 | | | | |
| Men, 50 year + | 73 | | | | |
| Women, less than 50 year | 84 | | | | |
| Women, 50 year + | 106 | | | | |
| In addition, destination Oslo rush | 32 | | | | |
| **Car passenger (CP):** | | | | | |
| All car passengers | | | 61 | 55 | |
| Men | 92 | 64 | | | 198 |
| Women | 154 | 147 | | | 253 |
| In addition, large city | | 58 | | | |
| In addition, weekend | | | -19 | | -84(m), -77(f) |
| **Public transport (PT):** | | | | | |
| Invehicle time | | | 33 | 28 | 92 |
| Invehicle time, men | 38 | 56 | | | |
| Invehicle time, women | 48 | 125 | | | |
| In addition, weekend | | | | | -32 |
| Access/egress time | 43 | 168 | 61 | 57 | 133 |
| Waiting time, when waiting 5 minutes | 60 | 369 | 70 | 84 | 124 |
| Waiting time, when waiting 30 minutes | 24 | 150 | 29 | 34 | 51 |
| Waiting time, when waiting 60 minutes | 17 | 106 | 20 | 24 | 36 |
| Transfer (NOK per transfer) | | 39 | | 16 | |

In all five models some of the values of time are higher for women than for men, which we believe can be related to women having a tighter time schedule than men, a stronger perception of travel time as discomfort, and a travel pattern with significant shorter trips than men. Generally shorter travel distances lead to higher value of time (separate models, or also separate time and cost variables, for men and women might lead to different results). In the business model, where the difference in value of time is largest, the value of time is nearly twice as high for women than for men. This can to some extent be explained by the fact that women and men make different types of business trips. Men are more frequently employed in the private sector, while women more frequently are employed by the local authorities or other public institutions (where business trips often are short, with destinations inside the boundary of the municipality). This leads to a travel pattern with longer business trips for men than for women, which in turn results in lower value of time for men.

School trips (trips to and from school) are excluded from the data in the estimation of the mode/destination choice models. The reason is that the public transport network/lines for school trips (school buses etc) are not coded in the network model. These trips are dealt with quite schematic, with gravity models distributing children and young people to schools and universities. It is not estimated any mode choice model for these trips.

# Models for choice of travel frequency

For each travel purpose (including school trips), the *frequency models* produce the expected total number of visits as a function of age, gender and other background variables, together with logsums taken from the mode/destination choice models. The data for estimation are the National Travel Survey 2001, with supplementary information on the number of visits the registration day for each of the travel purposes.

The concept is based on weighted logit models that take care of the distribution of the number of visits for *different travel purposes*, combined with hurdle-poisson models that simultaneously give the *total number of visits*. Poisson models are often used when we want to know the distribution of something (1,2,3,4,5,….), while hurdle-poisson is used when "0" is included in the distribution. The hurdle-poisson model then gives the expected total number of visits for the 6 travel purposes, while the distribution on purposes is given by the logit model.

We first tried to estimate a joint model for the entire population, but this resulted in a huge amount of segments with dummy variables for different age groups. Instead we estimated independent models for 5 different age groups (age 13-24, 25-34, 35-54, 55-66 and 67+). There are large differences in the split on travel purposes between the age groups (table 2), which were one of the reasons for needing different models.

*Table 2. Distribution of visits on different travel purposes for each age group/model. Week day, all of Norway.*

| Purpose/age | 13-24 | 25-34 | 35-54 | 55-66 | 67+ | Total |
|---|---|---|---|---|---|---|
| Work | 12 % | 26 % | 31 % | 28 % | 3 % | 24 % |
| Business | 3 % | 9 % | 11 % | 9 % | 1 % | 8 % |
| Shopping/service | 19 % | 24 % | 24 % | 29 % | 53 % | 26 % |
| Private visitis | 15 % | 10 % | 7 % | 11 % | 14 % | 10 % |
| Other private trips | 26 % | 29 % | 27 % | 23 % | 30 % | 27 % |
| School | 24 % | 3 % | 1 % | 0 % | 0 % | 6 % |
| Total | 100 % | 100 % | 100 % | 100 % | 100 % | 100 % |

The model for each age group is specified by one utility function for each travel purpose, formulated with alternative-specific constants and variables describing sex, age, family type, type of residence, and a logsum from the corresponding mode/destination model (which implies that the travel frequency is influenced by the accessibility described by the mode/destination models, i.e. level of service, attractiveness of destinations, and car availability). In addition we had to introduce some regional dummy variables.

# Procedure for calculation of "intermediate" trips

The models for mode/destination are based on round tours with only one main destination, while the models for travel frequency gives the total number of visits made. If the results from these models were used directly, the model system will produce too many trips (because some of the visits in practice are "intermediate" trips, and not a round trip). To correct for this we have made a *procedure that takes care of round trips with intermediate destinations*. The procedure uses the information from the mode/destination models and the frequency models in such a way that the system produces the correct number of outbound trips, intermediate trips and return trips, based on matrices of "transfer probabilities". "Transfer Probabilities" reflect the probability for a visit of a certain purpose to be followed by either a return trip or a new visit, and in the latter case which purpose.

The matrices of “transfer probabilities” are calculated based on data from The National Travel Survey 2001. For all trips that are not return trips, the purpose of the given trip and the next trip is registered. All such trip pairs are counted and put in a table, and the probabilities are calculated based on this.

# Results

A run of the model results in tour matrices for simple tour/return trips, together with matrices for outbound trips, intermediate trips and return trips for tours with more than one destination. For each travel purpose there are separate matrices for each of the modes. The matrices can then be used to compare the transport situation in different scenarios for a given year, to forecast traffic growth for different modes in a future year etc. If we want to know the distribution of the traffic on road links and on different public modes (the demand model calculates public transport as only *one* mode), the matrices have to be read into a network model. The trips will then be assigned on links and routes based on the algorithms of the network model.

At the moment the models for the first regions have been delivered to the clients, and the local road administrations are evaluating the results at the matrix level (comparing number of trips between different areas with traffic counts at road and other modes etc). At the same time SINTEF (a research centre in Norway) are working on the interface for the complete model (generation of LoS-data, running the demand model, assignment in the network model and presentation of results). The calibration of the model at network level (links) will be done by the local road administrations during the spring 2005.

# Bibliography

Ben-Akiva, M.E. and Lerman, R. (1985). *Discrete choice analysis: Theory and application to travel demand.* The MIT Press, The Massachussets Institute of Technology. ISBN 0-262-02217-6.

Hamre, T. N. (2002). *NTM 5 - The Norwegian National Transport Model, version 5.* Oslo, Institute of Transport Economics. TØI report 555/2002. ISBN 82-480-0241-1. (In Norwegian, with English summary.)

Hamre, T. N., Rekdal, J. and Larsen, O. I. (2002). *Development of the Norwegian National Transport Model in phase 5. Part B: Estimation of models*. Oslo, Institute of Transport Economics. TØI report 606/2002. ISBN 82-480-0297-7. (In Norwegian, with English summary.)

Larsen, O. I. (2004). *Estimation of trip generation models*. Molde, Molde University. Working paper. (In Norwegian)

Larsen, O. I. and Rekdal, J (2004). *Treating seasonal tickets for public transport correctly in estimation and application of mode/destination choice models*. Molde, Molde University.

(Paper presented at Tralog 2004).

Larsen, O. I (2004). *School trips.* Molde, Molde University. Working paper. (In Norwegian)

Larsen, O. I. (2003). *Estimating independent and simultaneous trip frequency models for all travel purposes with combined Logit/Poisson.* Molde, Molde University. Working paper 2003:7. ISBN 82-7962-040-0.

Larsen, O. I. (2003). *Estimation of models for choice of mode and destination for work trips.* Molde, Molde University. Working paper 2003:9. ISBN 82-7962-042-7. (In Norwegian)

Larsen, O. I. (2002). *Trip chains and frequency models, illustrated with data from Hordaland.* Molde, Molde University. Working paper M 0213. (In Norwegian)

Larsen, O. I. (2002). *Implementation of a new long distance model – segmentation according to license holding and car availability.* Oslo, Institute of Transport Economics. Working paper PT/1555/02. (In Norwegian)

Madslien, A (2005). *Estimation of models for choice of mode and destination for private visits.* Oslo, Institute of Transport Economics. Working paper 1705/2005. (In Norwegian)

Madslien, A (2005). *Estimation of models for choice of mode and destination for shopping trips.* Oslo, Institute of Transport Economics. Working paper 1706/2005. (In Norwegian)

Rekdal, J (2004). *Segmentation models for car ownership and license holding. Models for use in regional transport models based on basic statistical units.* Molde, Molde University. Report 0410. ISBN 82-7830-070-4. (In Norwegian)

Rekdal, J. and Madslien, A. (2004). *Development of regional (short distance) transport models in Norway.* Molde, Molde University and Oslo, Institute of Transport Economics. (Paper presented at Aalborg Trafikdage, 2004. In Norwegian)

Rekdal, J (2003). *Estimation of models for choice of mode and destination for business trips.* Molde, Molde University. Working paper 2004:8. (In Norwegian)

Rekdal, J, Larsen, O. I., Hamre, T. N. (2002). *Development of first generation short distance regional models for private travel in Norway*. Work programme. Oslo, Institute of Transport Economics. Working paper PT/1580/2002. (In Norwegian)