Application of Models based on Stated and Revealed Preference Data for Forecasting Danish International Freight Transport

Goran Jovicic, TetraPlan AS

June 1998

1. Introduction

TetraPlan A/S has developed a mode choice model for Danish international freight transport, under the umbrella of TRANSFORSK '95 projects. The objective of the model is to forecast the future demands for rail and sea transports relative to lorry transport, when supply variables change. To achieve the above goal a number of ideas are built in the model structure. These ideas are listed below:

- To combine data in the model structure which not only different in their nature, i.e. revealed and stated preference data, but also data collected in a number of independent domestic freight projects,

- To include as many as possible variables in the model structure, which can be found in the revealed preference data, and measure their importance in the forecasts relative to transport costs and transport time, and

- To find a method of how variables found only in the stated preference data can be used for the forecasting purposes.

The aim of the paper is to discuss the points above in more details. Chapter 2 describes the estimation procedure from a number of aspects. Estimation results and some forecasting results are given in chapters 3 and 4, respectively. The final remarks are given in the last chapter.

2. Estimation Procedure

2.1 Model Structure

Demand modelling methods in the 1990's have used both Stated Preference (SP) and Revealed Preference (RP) data simultaneously to estimate the model parameters, exploiting the strengths of each type of data (Bradley and Daly, 1992). These methods have improved significantly the reliability of models based on SP data. Further problems, however, have been identified with the application of these models, such as representativeness of the data samples and the possibility of using the variables estimated solely by the stated preference data in the forecasting context.

A two-step estimation procedure has therefore been developed to deal with these problems. In the first step of this procedure, both RP and SP data are used simultaneously to estimate trade-off parameters, applying a scaling factor to the SP parameters to account for the different variance between the RP and SP data. In the second step, the RP data only is used to estimate alternative-specific constants for the RP alternatives, constraining the values of the other trade-off parameters to be equal to those estimated jointly in the first step.

The latest Storebælt and Femerbælt passenger models follow this methodology. These two models deal, however, with an extra problem of treatment of the new alternatives in the forecasts (Daly, Rohr and Jovicic, 1998).

This model can be described as a hierarchical multinomial logit model based on the combination of RP and SP data. Transport modes included in the model structure are lorry transport (described further as Road Transport (RT)), conventional rail together with the combined road-rail transport (described further as Combined Transport (CT)), and sea transport (ST). Based on the experience from the previous goods transport projects in the country and abroad, CT and ST alternatives are grouped together vs. the RT alternative, describing therefore a higher cross-elasticity between the first two modes in relation to RT. We postulated therefore that the shippers distinguish first of all between RT and non-RT modes and after that among the non-RT modes they distinguish between CT and ST modes.

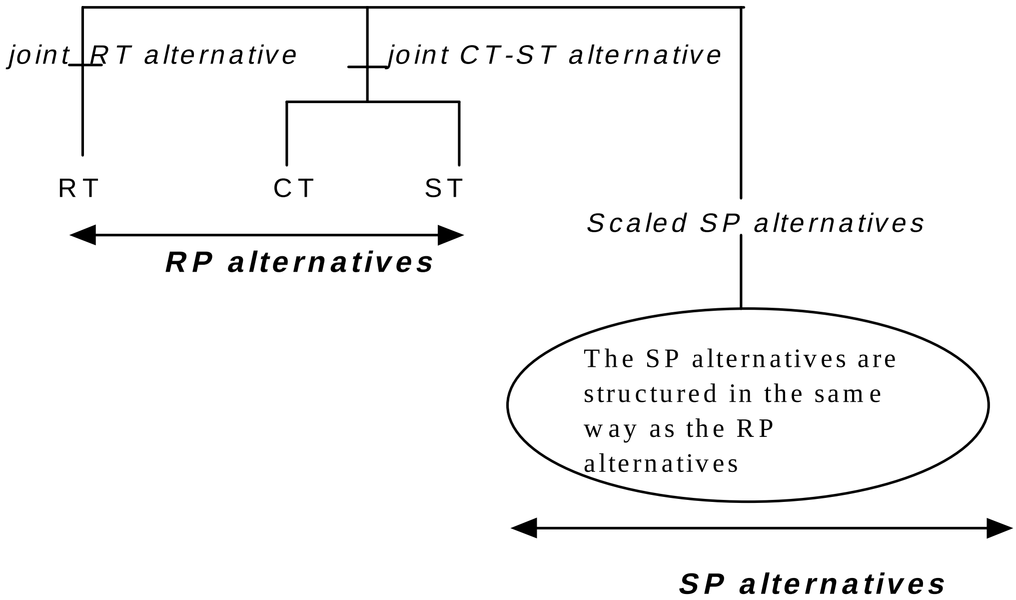
Due to the differences in the error standard variance between the RP and SP data, the stated preference data are finally scaled to the RP data. It has been taken care of differences between the sources of data (see paragraph 2.2). The across-mode SP data are both scaled to the RP data and structured in the way that CT and ST modes are grouped vs. the RT alternative. A simplified model structure is shown in figure 1.

A structural parameter (usually called the theta parameter) is present in a dummy *joint CT-ST alternative* to reflect the higher cross-elasticity of choosing between these two modes relative to the choice between the RT and non-RT alternatives. If this parameter is found to be significantly different from 1, the model structure presented in figure 1 will be justified. It is important to note that the same parameter is also present for the RT alternative (i.e., dummy *joint RT alternative*) to ensure consistency in the scale of the utility functions.

When scaling SP data to RP data it is recommended that data from each SP game, for each data source, have its own scaling parameter, reflecting the possible differences among them. This scaling parameter can have any positive value. If the scaling parameter is found to be between 0 and 1 (which is typical for across mode SP data), then SP responses have more unexplained variance than RP data. Opposite to that, if the scaling parameter is found to be greater than 1 (which is typical for within mode SP data), then SP responses have less unexplained variance than RP data.

There are seven scaling parameters defined in the model originating from seven SP games.

Figure 1: The model structure



2.2 Model data

Data applied in the model originates from three different projects:

- The Nordic-Link project,
- Parameters of Quality of Goods Transport project, and
- Potential of Sea Transport project.

The RP and SP data have been collected in all three projects. The data in the first two projects are completed in the period 1996-1997, while the third project is from 1993. All the data are structured in the model as shown on figure 1.

The Nordic-Link project data consist of RP and SP data. The SP data are further divided in three sets, one within mode and two across-mode SP-games data.

Parameters of Quality of Goods Transport project data consist of RP and SP data, too. The SP data are further divided between within mode and across-mode SP-games data. The within mode SP data are completed through 3 SP games, for the three modes separately. The reason for such a high number of SP games is that the effect of seven variables on the respondents' transport behaviour has been investigated. These variables are listed in paragraph 2.3. The across-mode SP data originates from one SP game where the respondents had a choice between the original transport mode and the first best alternative.

Potential of Sea Transport project data consist of RP and SP data. The SP data are further divided in three sets, all the within mode SP-games data for the three transport modes.

All together, the data applied in the model consist of 11337 observations in total, i.e., both RP and SP observations. Number of the RP observations alone is 1012, which represents about 11 per cent of the total. This is regarded as a good balance between the two sets of data, i.e. each respondent gave answers to about 10 hypothetical (SP) choice situations. A less desirable characteristic of the sample is that the great majority of the respondents are original users of road transport, i.e. there are 847 RT respondents in the sample. This equals to about 85 per cent of the sample. 90 respondents are originally the CT users (about 9% of the sample), and the rest of 75 respondents (about 7% of the sample), in the sample of 1012 respondents, used originally sea transport.

2.3 Model Variables

Transportrådet published last year a report, which described quality in goods transport from a number of different angles (notat nr. 97–02). One of the approaches in the project was to measure the relative importance of a number of variables in the present transport behaviour of RT, CT and ST users. A methodology of SP analysis has been employed for this purpose (Jovicic, 1997). Apart from the usually applied variables in this type of projects, i.e. *transport cost* and *time*, *risk of damage* and *delay* and *frequency of travelling*, two new variables are applied, namely *flexibility* and *information system*. *Information system*, for example, as the final results showed, proved to be rather important for the present users of rail transport. This model couldn't, however, say anything about the importance of these two variables in the mode choice among the sampled respondents.

The present mode choice model includes 7 variables, i.e. *transport cost* and *time*, *risk of damage* and *delay*, *transport frequency*, *information system* and *flexibility*. The model can be therefore seen as a continuation of the project described above. Table 1 lists the variables used in the project. In column 2 is given a description of the variables' units. The source of the variables is given in column 3.

The estimation procedure of the variables above follows the model structure shown in figure 1. In step 1 of the estimation all the data (i.e., RP and SP data) are used simultaneously to determinate the trade-off parameters. In the second step, the mode constants are adjusted to the population's size, i.e. the model represents the total traffic to and from Denmark.

2.4 Model Segmentation

From the practical point of view, model segments should have different values of travel time (VOT). In the previous two Transportrådet's goods transport projects (notat 95-02 and notat 95-03) the model segments applied in the model structure were:
- Low-value shipments: shipments with values till 20 DKK/kg, and
- High-value shipments: shipments with values greater than 20 DKK/kg.

The problem with these two models, in relation to their segments, is that it is difficult to calibrate them. It has been therefore suggested to find some other ways of calibrating the present model. Two segmentation variables have been pointed out. These are *commodity groups* and *travel distance*.

Table 1 – Model Variables

Variables	**Variables' Units**	**Variables' Source**
Transport Cost	Door-to-door transport cost in DKK	RP and SP data
Transport Time	Door-to-door transport time in hours	RP and SP data
Risk of Damage	Per mille of the risk of the shipment being damaged at the destination	RP and SP data
Risk of Delay	Per cent of the risk of the shipment being delayed at the destination	RP and SP data
Transport Frequency	Number of weekly departures for rail and see transports	RP and SP data
Information System	A descriptive variable which has two levels: Level 1: Low level of information Level 2: High level of information	SP data
Flexibility	A descriptive variable which has two levels: Level 1: Low level of flexibility Level 2: High level of flexibility	SP data

The Nordic Link project had defined 8 commodity groups. *Parameters of Quality of Goods Transport* and *The Potential of Sea Transport* projects had defined 11 commodity groups. After aggregating these two commodity definitions, the mean values per kilogram were determinate for each commodity group. These values were used as measurement units for defining the model segments in relation to commodity groups. Table 2 lists commodity groups joined in the model segments. Note that the bulk commodities are not included in the model because they are unlikely to shift the mode.

The commodity group named as *other* was used by 25% of the respondents. This commodity group couldn't be therefore avoided in the estimation procedure. Its mean value per kilogram suggested that this commodity group should be included in segment 2. Observations referring to perishable vegetable- and animal-products, as being highly dependent of travel time, are also included in segment 2.

It is expected that value of travel time declines with travel distance due to the modal shares in different distance intervals. The data have allowed us to define 3 segments in this model:

- Segment 1: Observations with the travel distance up till 600 km,
- Segment 2: Observations with the travel distance between 600 and 1200 km, and
- Segment 3: Observations with travel distance above 1200 km.

Due to the space restrictions, the remaining part of the paper only deals with the model segmented by commodity-groups.

Table 2 – Model Segments Defined in relation to Commodity Groups

Segment 1 – low value commodity groups	**Segment 2 – high value commodity groups**
1. Vegetable Products	1. Chemical Products
2. Animal Products	2. Machinery
3. Beverages	3. Electrical Products
4. Feeding Stuffs	4. Textile and clothing
5. Metal Products	5. Other Products
6. Paper and Wood Products	

3. Estimation Results

The trade-off values together with the t-values for both low value and high value commodity groups are shown in table 3.

Table 3 – Estimation Results for the Commodity Groups Models

Parameters	**Model for Low Value Commodities**		**Model for High Value Commodities**	
	Estimates	t-vales	Estimates	t-values
Travel Cost	-0.0003790	-3.50	-0.0005095	-3.90
Travel Time	-0.0053360	-3.30	-0.0234400	-4.70
Risk of Damage	-0.0260900	-2.70	-0.0640000	-4.00
Risk of Delay	-0.0359400	-3.00	-0.0708300	-3.50
Travel Frequency	+0.0590800	+3.10	+0.1356000	+2.70
Flexibility	+0.1447000	+2.40	+0.1156000	+1.50
Information System	+0.2092000	+2.40	+0.0690000	+0.70

All parameters are estimated with the correct sign. T-values are also significantly higher than 1.96, which indicates a 95% statistical confidence that the parameters' values are different from zero, i.e. the parameters significantly contribute the model fit to the observed data. The only exception to that are parameters of *flexibility* and *information system* variables in the model for high value commodities. The reason why they are maintained in the model structure for the forecasting purposes is that their estimates have the correct sign.

The best practical check of the model estimation results is to make a relative relationship between the parameters, e.g. to establish monetary value of time and other parameters. This is done by dividing the estimates for the time and other variables with the travel cost parameter. The results of this test are presented in table 4.

Table 4 – Monetary Values of the Applied Variables

Value of	**Unit**	**Model for Low Value Commodities**	**Model for High Value Commodities**
Travel Time	DKK/hr	14.08	46.01
Risk of Damage	DKK/per mille of damage	68.84	125.61
Risk of Delay	DKK/per cent of delay	94.83	139.02
Travel Frequency	DKK/no. of weekly departures	155.88	266.14
Flexibility	DKK/per level of service	381.79	226.89
Information System	DKK/per level of service	551.98	135.43

The producers of high value commodities in the sample are willing to pay about 3 times more for saving one hour of transport time between origin and destination than the producers of low value commodities. Value of risk of damage for the high value commodities is about twice as high as the value for the low value commodities. Values of risk of delay and travel frequency are about 50 per cent and 70 per cent, respectively, higher for high value commodities than for low value commodities.

Flexibility and information system variables are valued higher by the producers of low value commodities than by the producers of high value commodities. From the statistical point of view we cannot trust fully the trade-off values of these two variables for high value commodity groups (see table 3). However, it is reasonable to postulate that both flexibility and information system is already on a high level among lorry operators (RT is the most commonly applied transport mode for this type of commodities) so that further improvements are not of high importance to respondents in the sample.

4. Adjustment to Observed Mode Constants and Forecasting Results

4.1 Adjustments to Observed Mode Constants

Because the base estimation uses unweighted data, the alternative-specific constants are not correct, i.e. they do not reproduce the observed alternative shares for the total transports between Denmark and abroad. These constants therefore need to be adjusted so that the observed mode shares in the base year are replicated. An automated procedure has been developed where the RP data are used to re-estimate the alternative specific constants for RT, CT and ST alternatives, while constraining the trade-off coefficients to the values found in the estimation process.

The model is calibrated after the observed Danish export/import volumes for the two commodity group-segments and three modes of transport in 1995 (statistical yearbook for 1996). Table 5 summarises the transport volumes, in millions of tons, applied in the calibration process.

Table 5 – Danish Export/Import per Transport Modes and Commodity Types in 1995, Millions of tons

Transport Modes	Low Value Commodities	High Value Commodities
Road Transport	7.747	8.224
Rail Transport	1.206	0.728
Sea Transport	7.226	5.185

The calibration phase is an iterative procedure where the correction factors to the alternative-specific constants, found in the estimation phase, are calculated using the following formula:

Correction Term = Old correction value + ln(95 observed value/model result)

Note that *'old correction term'* equals zero in the first iteration. The values of the alternative specific constants before and after the calibration phase are shown in table 6.

Table 6 – Values of Alternative Specific Constants before and after the Calibration Phase

Transport Modes	Low Value Commodities		High Value Commodities	
	Before calibration	After calibration	Before calibration	After calibration
Road Transport	0	-0.10	0	-0.48
Rail Transport	-0.24	-1.11	-1.18	-1.97
Sea Transport	+0.003	+0.41	-0.92	+0.30

The size of alternative specific constants, both before and after the model calibration, shows that the model is very well described by the chosen variables, which makes a good base for the following forecasts.

One of the new aspects in the project is that two variables, found originally in the SP data only, namely *flexibility* and *information system*, are applied also in the forecasts. To do that it is necessary to attach values for these two variables for each of 1012 RP respondents before the calibration. We have therefore assumed that, for example, RT *information system* is better than CT and ST *information system*. The same goes for *flexibility*. Assuming that these two variables can take a value between 1 and 2 for each RP respondent, the RT *information system* and *flexibility* are set to 1.6 both, while the values of these two variables for CT and ST are set to 1 (the lowest possible level).

4.2 Forecasting Results

Forecasts consist of model runs where one, or more, variables change values relative to the original values. The forecasting results for one of these scenarios, where the road transport cost is increased for 10 per cent, are shown in figure 2.

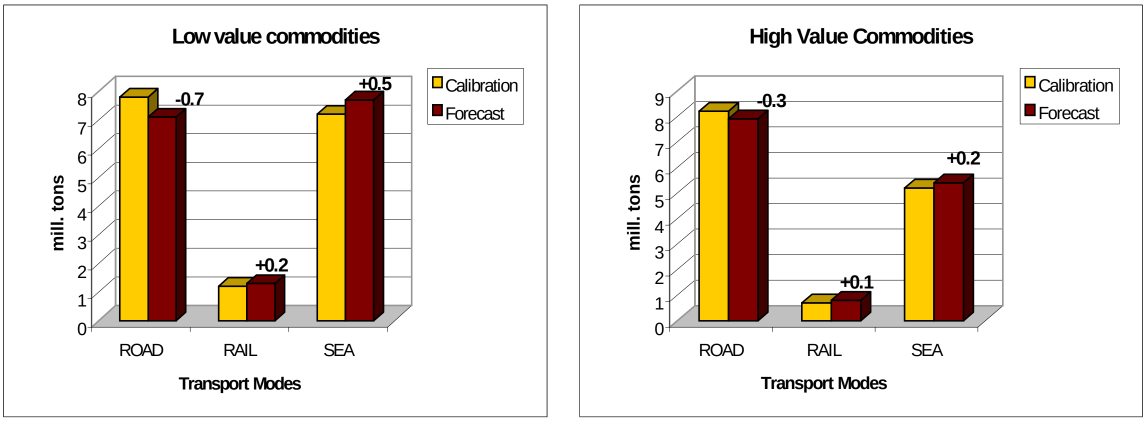
Transport cost is, as expected, more important for low value commodities than for high value commodities. This is why 700.000 tons of road transport goods will transfer from RT to CT and ST for low value commodities in this scenario, while “only” 300.000 tons will shift away from RT for the high value commodities.

The forecasting results for different scenarios lead to the following two conclusions:

1. Among 7 variables applied in the model structure (see table 1) it is *transport cost* that has the greatest importance in the mode choice in both model segments. *Transport time* variable is less import than *transport cost*, but more important than *risk of damage*, *risk of delay* and *transport frequency*. *Flexibility* and *information system* variables are of least importance in the mode choice in both model segments.

2. The biggest share of the RT tons, according to the forecasting results, shifts to ST and the smallest share shifts to CT. This points out that ST is more competitive to RT than CT in the Danish export/import. This is in line with the conclusions from a domestic goods transport project described by Henriques, 1998.

Figure 2 – Forecasting Results for the Increased RT-cost for 10 %



5. Conclusions

A hierarchical logit mode choice model for Danish international goods transport has been built in this project. The project gives some practical solutions for model work with different sources of data, both RP and SP data and data from different project-sources. It has been further shown in the paper how variables, originally applied in the SP data only, can be used in the forecasts.

The sample of 1012 respondents gave about 11000 observations. The model presented in the paper consists of two segments, i.e. low value commodity groups and high value commodity groups.

All together, seven variables are applied in the model structure. Model validation (see table 4) shows that the chosen variables significantly contribute to the model fit to the observed data. That further means that the importance of the alternative specific constants is lowered to the least. This gives a good base for reasonable forecasts.

The forecasting results show that *transport cost* is the far most important variable when choosing a mode of transport at the project. *Transport time* is less important than *transport cost* but more important than the rest of the applied variables in the model structure. *Risk of damage* and *delay*, and *transport frequency* are of the secondary importance when Danish importers and exporters choose transport mode for their shipments. Finally, *information system* and *flexibility* are of the least importance among the seven variables.

The forecasts also show that sea transport is more competitive to lorry transport than rail transport in the Danish international goods transport.

References:

1. Bradley and Daly, 1992. 6th International Conference on Travel Behaviour, Quebec, Canada. Estimation of Logit Choice Models using mixed Stated Preference and Revealed Preference Information

2. Daly, Rohr and Jovicic, 1998. 8th World Conference on Transport Research, Antwerp, Belgian. Application of Models based on Stated and Revealed Preference Data for Forecasting Passenger Traffic between East and West Denmark

3. Transportrådet, 1997. Notat nr. 97-02. Godstransport og Kvalitet

4. Jovicic, 1997. Notat nr. 97-02. Study on Importance of some Quality Parameters in Long-distance Goods Transport – Application of Stated Preference Analysis in the Study

5. Transportrådet, 1995. Notat nr- 95-02. Potentiale for søtransport mellem Danmark og kontinent

6. Transportrådet, 1995. Notat nr- 95-03. Valg af transportmiddel i international goodstransport – en stated preference analyse

7. Henriques, 1998. Trafikdage på Aalborg Universitet 1998. Transportkæders miljø og omkostninger