SENSITIVITY OF VARIABLE DEFINITIONS IN SP-ANALYSES - AN EMPIRICAL STUDY OF CAR-USERS' EVALUATION OF LENGTH, COST AND TIME

Otto Anker Nielsen and Majken Vildrik Sørensen
Centre for Traffic and Transport (CTT), Building 115, st.tv. Bygningstorvet, Technical University of Denmark (DTU), 2800 Lyngby, Denmark, Fax: (+45) 45 93 64 12
oan@ctt.dtu.dk, mvi@ctt.dtu.dk

Abstract

The paper investigates how the definition of cost and length variables in an Stated Preference Experiment (SP) affects the results. This was done by asking questions in different ways to the same persons during an SP-experiment, running different experiments on different persons, and comparing these with results from a Revealed Preference experiment (RP) carried out on the same persons after the SP. The results pinpoint that coefficients, derived elasticities and value of times are highly sensible towards the definition of variables in SP. The respondents had difficulties in estimating the cost of driving per km., which was often inconsistent with reality, as well as with the cost-components, the respondents claimed they had included in their cost estimates. The respondents had also a tendency to overstate their responses to changes in attributes in the SP, as they in the following field experiment (RP) changed behaviour to less extent. The Value of Times (VoT) differed with more than a factor 2 depending on how costs had been defined in the questionnaire. The type of model did affect the results as well; especially Mixed Logit models improved fit significantly in the SP-experiment compared to Multinomial or Nested Logit. The improvement in the RP by using random coefficients was even larger.

Keywords: Stated preference; SP; Mixed logit; Design of SP; Revealed preference
Topic area: D1 Passenger Transport Demand Modelling

1. Introduction

The aim of the paper is to describe how the definition of cost and length variables in a Stated Preference Experiment (SP) affects the results. This was examined 1) by testing different definitions in a systematic way within an applied Stated Preference Experiment, 2) by examining the sensibility of different estimation methods and model assumptions, among others cost-definitions, NL versus Mixed Logit (ML) and different assumptions on the random coefficients in the ML, and 3) by comparison with the results from a RP-survey where the same persons' cars were followed over a 4-6 month period. Some of the initial results were presented in Nielsen & Jovicic (2003).

It was hereby possible to test different variable definitions in the SP, and compare the results with the RP. The SP was carried out prior to the RP to secure that the experiences in the RP did not affect the results of the SP.

Section 2 of the paper presents the experimental design and section 3 presents some more general results. Section 4 presents and discusses the SP-model and section 5 the RP model with focus on route choice, following by a comparison of the two in section 6; conclusions are given in section 7.

2. Experimental design

AKTA is the Danish part of the EU-project PROGRESS, where road pricing is investigated by various approaches in 8 European cities. The behaviour of the participants was followed by GPS with and without a road-pricing system. AKTA is unique as each car was followed over a fairly long period, which provides more information on interpersonal variation than in most RP-datasets.

The purpose of AKTA was to test whether road use taxes will change travel behaviour. Hence, the city of Copenhagen was equipped with virtual cordon rings and prizing zones. The voluntary test drivers were equipped with a GPS-based vehicle position system, making it possible for them to read the virtual prizing systems on a display. The cars' movements were logged in the system, and a road price cost calculated for every trip. Different pricing schemes including a pure control period was followed in a first and second 8-week period in the first round (the first 200 participants), and 10-week periods in the second round (another 200). In the third round all 100 participants were subject to a 10-12 week control period followed by the same pricing scenario for all for 12-16 weeks. At the end of each round, the test drivers were paid according to an estimate based on the difference in behaviour between the two periods.

The 300 participants from second and third rounds participated in a SP-experiment before the road pricing experiment. This survey is the basis for the work presented in the present paper.

2.1. Survey design

The participants were distributed after a full factorial design among income groups, commuting patterns (residence and work) and pricing schemes. All participants belonged to one-car families and resided and/or had their workplace within the road pricing area (the geography of Copenhagen eliminates through traffic for Commuting). All participants had a daily need of transport.

The participants completed a questionnaire before the experiment, and another plus a telephone interview after; among other things to test whether they changed attitudes.

For comparison, a telephone interview with 1,015 respondents living in the road pricing area was carried out to investigate the general populations awareness and attitudes towards road pricing as well as to control for possible sample bias (i.e. that the participants in the experiments were representative of the population).

2.2. Pricing schemes

Two pricing schemes were zone-based with four different pricing levels per km. with the most expensive in the inner city, and the cheapest in the suburbs (figure 1). The third scheme was a cordon based system with payment for zone border crossings. The pricing was higher at peak than off-peak hours for all scenarios.

The GPS-device dynamically calculated the pricing level, and the participant could for a given trip see the pricing level (zone), be noticed on zone shift (cordon), and read the cumulated cost of the trip.

The coordinates were logged each second. They were then imported into a digital map (by GIS-technology) after the field experiment was completed and related (map-matched) to roads, junctions and origin/destinations of trips.

A control period was introduced to obtain information on participants usual driving pattern; i.e. time of travel, trip length, usual number of destinations, number of stops on a round trip etc. In the control period there were no pricing for road use and the display was turned off such that the participants were not expected to change behaviour at this time.

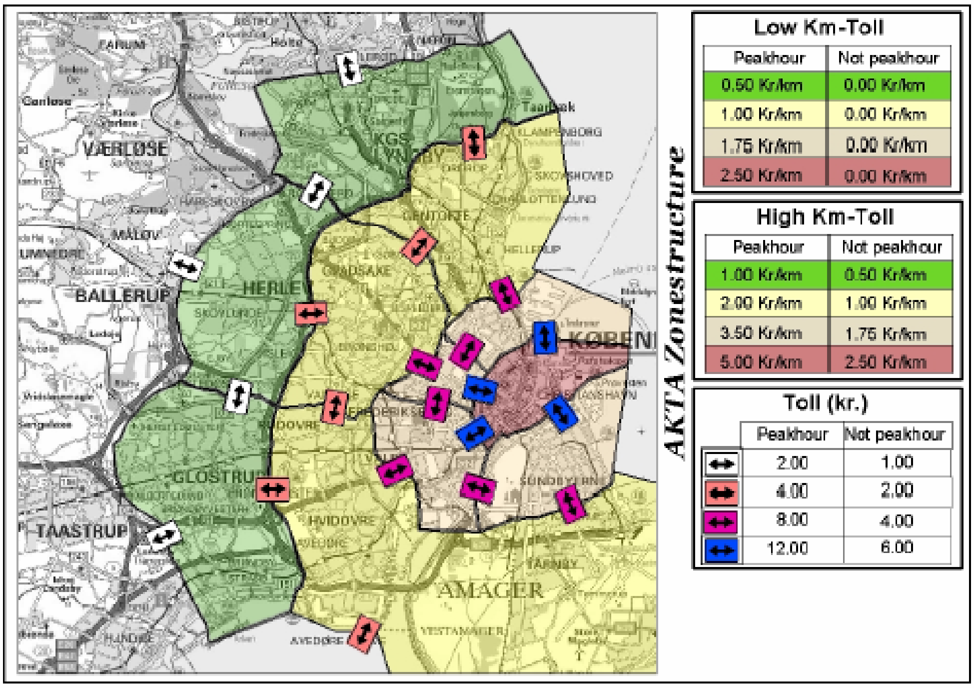

Figure 1 The different road pricing and toll schemes.

2.3. SP-Design

The Technical University of Denmark and the Danish Transport Research Institute (DTF) carried out a research project concerning Stated Preference (SP) methods in parallel with the main AKTA experiment. The purpose was to evaluate SP as a method, since AKTA provides an excellent RP-data survey to compare with. 200 car drivers from the second and 100 from the third round were interviewed while they waited for the GPS-unit to be installed in their car at the garage. Hereby the SP data was collected prior to start of the trial period and further. This reduced the cost of the SP survey and ensured that most of the 300 participated in the SP-experiment. 279 interviews were successfully completed and processed resulting in 3,388 SP records passed the data quality control.

The SP questionnaire collected information on a ‘typical trip’ e.g. the journey to work, for which the respondent was expected to have a sound knowledge. The chosen trip was described by including origin and destination addresses, departure and arrival times, and travel purpose. If the respondent completed an extra activity on the way (e.g. shopping, visiting bank) these activities were also notified. The departure time defined if the trip should be understood as the ‘peak’ or ‘out-of-peak’ trip.

The experiments considered the trade off between travel time and cost; choice of time of day (peak vs. off peak), and road pricing scenarios. For half of the respondents the cost of travel was in monetary units as usual in most value of time SP's; for the other half the cost was measured in distance (kilometres).

The pricing levels in the SP experiment were the same as the scheme the specific respondent would face in the main AKTA experiment. The SP-experiment was carried out before the RP to avoid that answers influenced by the experience. Each respondent then followed 3 SP-experiments before answering questions on socio-economy:

- • The first SP-experiment focused on the Value of Time (VoT). Traditional experiments trade of time towards cost. However, some prior experience in Denmark (Nielsen *et al*, 2002), suggest that some car-users are not aware of the cost, and that most travellers considered only marginal costs (fuel). To investigate this issue further, some of the respondents were asked to trade of time versus cost (referred to as SP1a in the following), and some time versus length (SP1b).
- • The second SP-experiment focused on time-of-day (ToD) decisions and congestion[1]. The trade off context in this respect was travel cost; travel time, extra time due to congestion and time-of-day (peak hours or not). Travellers in the peak hours may choose to travel outside the peak due to lower cost or congestion. While travellers presently travelling outside the peak may choose to travel in the peak hours, if additional costs make this faster than today. Also this experiment had a trade-off of time versus cost (SP2a) or length (SP2b).
- • The third SP-experiment evaluated the choice between an existing trip, where road pricing had been added, and a possible alternative route, which was found together with the respondent. The trade-off was then investigated between cost, pricing, free time, and congestion time.

3. General results

The following section presents some general results of AKTA; modelling results from the SP and RP-experiments are presented in section 4 and 5 respectively.

3.1. Participants in the AKTA experiment

The difficulties of recruiting a sufficient number of participants turned out to be far more complicated than anticipated; a total of 25,000 people had to be contacted in order to get a sample of 500 one-car households distributed over three rounds of experiments. Fortunately, during the experiment the drop-out rate was very low though.

Participants were presented for three ordinary questionnaires and one combined RP and SP interviewer assisted questionnaire. The households' primary car driver generally answered all these. The three rounds were different in set-up such that drivers were tested under different road pricing schemes; the rounds were undertaken at different times of the year, different lengths of the test period. Only participants in round 2 and 3 were subject to the SP questionnaire.

An effort was made to appreciate that the participant were the households' primary car user and that this person responded to all inquiries including questionnaires concerning the experiment. However, two of the RP questionnaires were mail-back.

For 97% of the participants the primary user did participate in the surveys of these, 2 out of 3 were males; most were living in the central part of Copenhagen, hence affected by the road pricing schemes every day if introduced. 74 % used their car for the journey to work, and more than half actually had the possibility to change time of travel for that trip. The yearly demand for car usage was below 10,000 for 27% and 10-20,000 km for more than half of the participants.

---

[1] It is noted, that the definition of congestion time followed the definition in Nielsen et al (2002), i.e. that congestion time is the extra expected travel time caused by congestion compared to the situation without congestion. The respondents were asked how long time their usual trip took, and how long time they expected it would have taken without congestion. The difference was then interpreted as the extra time caused by congestion.

3.2. Results from the after survey

The participants were contacted by phone after the experiment to answer a few general questions about their behaviour and the experiment.

Table 1 show the stated saving strategies distributed on the different pricing schemes in the experiment. Although the procedures in the experiment had been explained very thoroughly both orally and written, many participants had misunderstood the design. Several participants stated e.g. that they changed behaviour in the control period, and further a few changed their stated behaviour in the control period only. In the designs with no control, one participant had only changed behaviour in the low km. charge period, but not in the high km. period. A total of 24 participants (13%) had more or less misunderstood the experiment. These participants clearly do not act utility maximising.

Table 1 Stated saving strategies for the first round with 201 cars (183 answers). Numbers and (percent). Note that the two periods could be in both order (e.g. high km. before control or visa/versa). But these have been joined in the table for simplicity. Shaded fields indicate illogical behaviour, bold desired behavioural effect (Nielsen & Jovicic, 2003).

Pricing levels	No saving	Saving 1st period only	Saving 2nd period only	Saving both periods	Total
Control + High km.	25 (60%)	2 (5%)	**8 (19%)**	7 (17%)	42
Control + Low km.	34 (63%)	1 (2%)	**12 (22%)**	7 (13%)	54
Control + Toll	6 (38%)	0	**4 (25%)**	6 (38%)	16
Low km. + High km.	9 (35%)	1 (4%)	5 (19%)	**11 (42%)**	26
Low km. + Toll	6 (35%)	0	0	**11 (65%)**	17
High km. + Toll	13 (46%)	0	1 (4%)	**14 (50%)**	28
Total	93 (51%)	5 (3%)	30 (%)	56 (31%)	183

Out of the participants that understood the experiment, some had chosen to change behaviour – other not: Some participants did not *believe* they had an alternative as the alternative was considered too inconvenient, or they did not consider the pricing high enough to make them change behaviour (their willingness to pay and value of time were too high). In the experiments with two charge periods, some participants did only change behaviour in the period with the highest charge level.

170 cars/households from the second round where interviewed, of which 95 tried to save money. Table 2 shows the changes in behaviour distributed on the different pricing strategies. A total of 17 participants had more or less misunderstood the experiment (10%). This was somewhat improved compared to the first round, but also disappointing, since extra care had been made to explain the experiment due to the experiences from the first round. This confirms that a fraction of car users behave accordingly irrational or not utility maximising even with a high level of information.

Note, that the results are exclusive random variation of trip patterns, since they only concerns the participants stated strategies.

Table 2 Saving strategies for the second 200 cars (172 answers). Symbology as in table 1.

Pricing levels	No saving	Saving 1st period only	Saving 2[nd] period only	Saving both periods	Total
Control + High km.	23 (69%)	0	**7 (21%)**	3 (9%)	33
Control + Low km.	13 (45%)	1 (3%)	**6 (21%)**	9 (31%)	29
Control + Toll	13 (48%)	0	**12 (44%)**	2 (8%)	27
Low km. + High km.	11 (39%)	2 (7%)	2 (7%)	**13 (46%)**	28
Low km. + Toll	12 (40%)	2 (7%)	4 (13%)	**12 (40%)**	30
High km. + Toll	7 (30%)	2 (9%)	1 (4%)	**13 (57%)**	23
Total	79 (40%)	7 (4%)	32 (19%)	52 (31%)	170

3.3. Awareness of travel distance and travel time

An important issue when attempting to measure the effect of road pricing schemes on user behaviour is the validity of the measured impact of the travel cost on car usage. Users may consider the cost of travel differently and thereby have a different trade off between road pricing charge and travel cost. To assert whether any such problems persist, the respondents were questioned on their awareness of actual costs for a particular trip. The questionnaire was coupled with an underlying assignment model whereby zone-to-zone distance of the trip could be calculated. Further the total cost was calculated by a constant cost per km. multiplied the distance.

The respondents were prompted to verify or correct the cost whether provided in monetary units or kilo-metres. The rate of accepted to suggested cost is shown in figure 2, where it can be seen that the respondents frequently corrected the suggested cost and length. Lengths were more often corrected than cost (43% corrected the length, 18 % the cost). This may be interpreted as the respondents are more aware of travel length than travel cost for a typical trip.

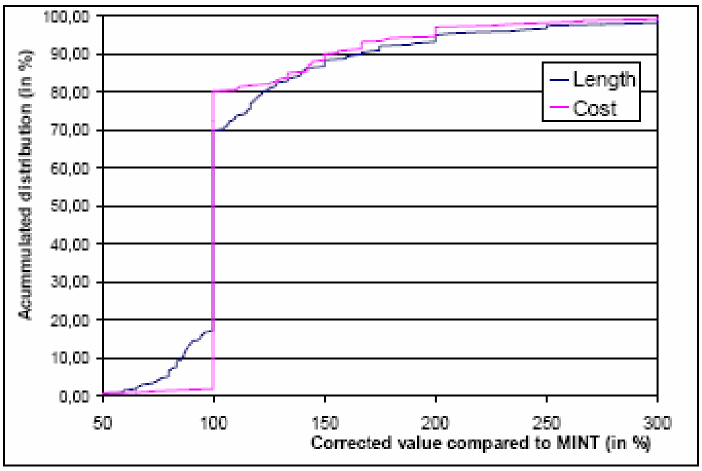

Figure 2 Respondents' correction of SP-estimates of cost and time based on zonal data.

Looking more into the corrections it appears that the correction of length had the same average size independent of trip length. This can be bound up with how the distances and costs were generated in the underlying assignment model where distances were measured between small zones (the error is related to the zonal connector in the from and to zones respectively); hence the aggregation error is independent of the distance between zones.
The respondents were then asked what they would normally consider the cost of driving per km. The typical answers were rounded values (0.5, 1, and 2 DKK were the most frequent answers), which can be seen in figure 3. The interval between 0.5 and 1 DKK equals reasonable well the marginal driving cost (about 45% of the answers) and values around 2 DKK equals average cost incl. capital cost, etc. (about 45% of the answers).

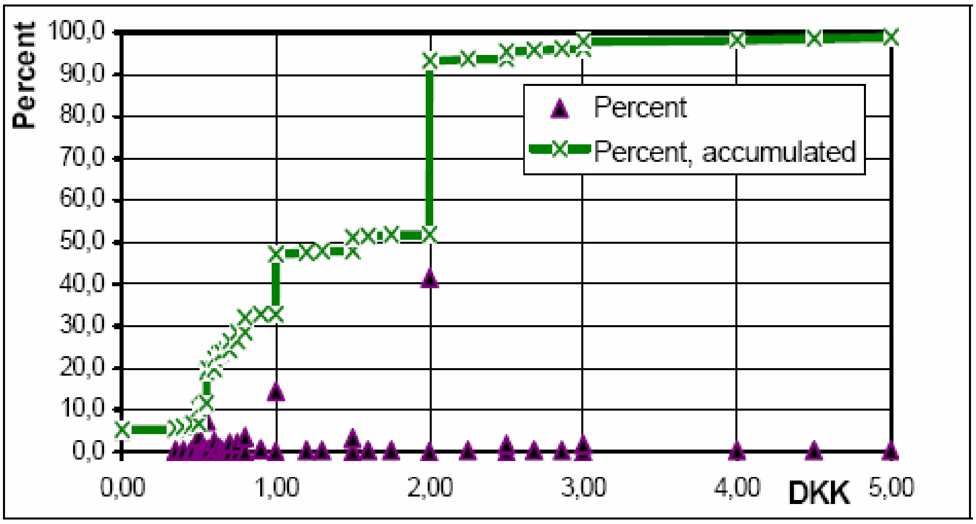

Figure 3 Accepted cost per km shown as cumulative and density distribution.

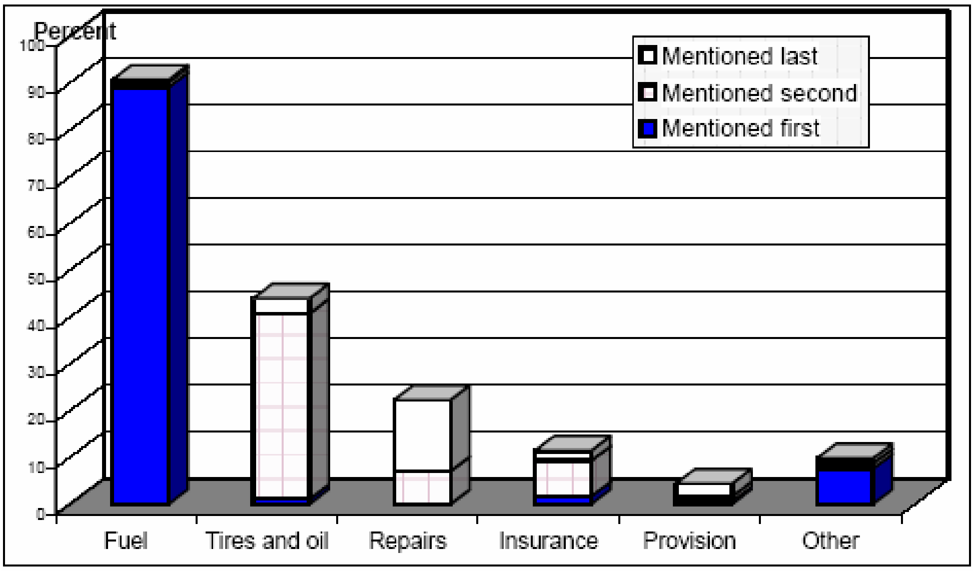

Figure 4 Participants stated components in their cost calculations.

As respondents may consider different cost components to be included in their cost per kilo-metre, respondents were asked for what the cost included (see figure 4). Only very few respondents claimed that fixed costs were included. There therefore there seem to be an inconsistency between the stated cost and the respondents' assumptions behind them. 6% of the participants even stated 0 DKK per km, since their cost of driving was paid for by other persons (employer, partner, friend, parents, etc.). About 4% of the answers in figure 3 were out of reasonable cost range (either too low or high cost per km.). About 48% stated accordingly a cost that was at a level which should include fixed costs, but only 10% stated that they have included fixed costs in their cost per km. estimates.

The answers from the respondents that stated their cost per km. could easily be compared to their *a priori* accepted length and cost for the specific trip (i.e. the ratio between the two). Figure 5 shows large disagreements between the answers provided by each person. As can be seen as many as 40% of the respondents (plus the 6% free of cost payment) answers deviated more than 100% dependent on the way the question was asked.

4. SP model estimation

As a first step a MNL model was formed on the data in order to eliminate variables that did not contribute to explain the variance in the data. The model included separate coefficients for cost and road pricing charge; travel time split into free flow time and delay related to congestion; a coefficient for distance for models where travel distance were presented instead of monetary cost; further dummies for time period (in or off-peak) and gender and alternative specific constants.

From this point several different strategies for model building were followed which included mixed logit formulation (normal and lognormal distributions), income dependent model formulations and combinations with mixed logit and income dependencies.

4.1. Model formulation (utility functions)

Several different utility functions was used in the estimation, which contained cost c (cost measured in monetary units or as length of the journey), road pricing, time t (free flow time and extra travel time due to congestion), socio-economic attributes s, and an error term ε assumed to be Identical Independent Gumbel Distributed *iid* over alternatives (1).

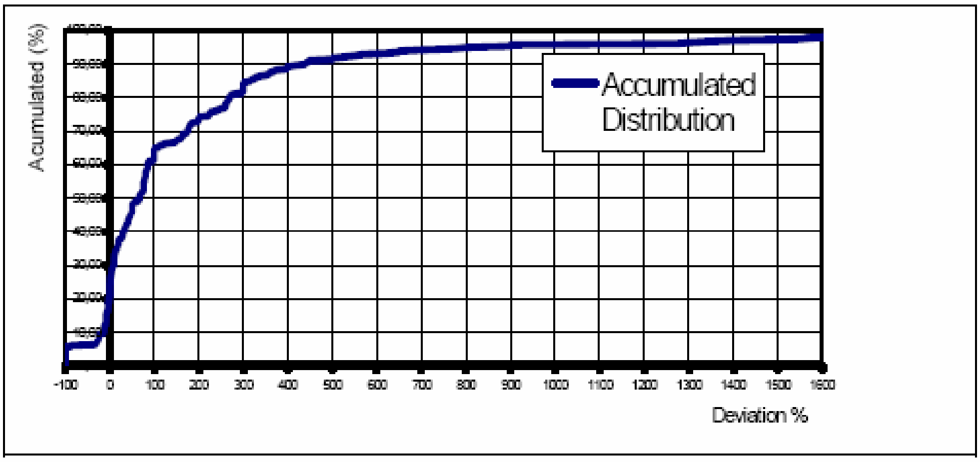

Figure 5 Comparison between accepted costs divided by accepted length (i.e. accepted total values) with stated cost per km. (%). Accumulated distribution.

In the second step of the modelling a mixed logit formulation was applied, whereby some of the coefficients were allowed to vary around their mean; this corresponds to substitution of coefficients β' in equation (2) for β in equation (1). Mixed models were estimated based on both normal and lognormal distributions assumed *a priori* as the mixing distributions. For distributed terms both the mean β and the standard deviance are presented (from these the parameters for the lognormal distribution can be calculated).

$$U_i = V_i + \varepsilon_R = \beta_c c + \beta_t t + \beta_{st} S + \varepsilon \qquad (1)$$

The best MNL and EC models are presented in table 3. Models 1 and 2 are MNL models while models 3, 4 and 5 are mixed logit formulations. In model 1 one cost coefficient corresponds to converting travel distances to cost by the fixed cost of driving (DKK/km) and evaluating the impact of road pricing to be of same scale as cost of driving. Model 2 deals with three cost coefficients; one related to driving costs, one related to driving distances (re-scaled into driving costs), and one related to road pricing. For the rescaling of distance to driving cost a fixed cost of 0.55 DKK/km was used.

$$\beta' = \beta + \xi \qquad (2)$$

Models 3 to 5 include one or more error components in its structure. The only difference between model 2 and 3 is that a random error was defined in model 3 connected to all time and cost coefficients (i.e. a very hypothetical situation, but a useful test of whether heterogeneity exist in data). A dramatic improvement is observed in model 3, showing a lot of taste variation. Further disaggregation of the error components in models 4 and 5 gave even better results as all proved to be significantly different from zero. The best model estimated is model 5 where error components were placed behind different cost coefficients, free flow travel time and congested travel time, i.e. six error components in total.

Table 3 Estimation results from the best MNL and EC models based on the SP-data

File	model 1		model 2		model 3		model 4		model 5	
Observations		3388		3388		3388		3388		3388
Final log (L)		-1662.6		-1645.7		-1571.1		-1561.3		-1538.1
D.O.F.		8		10		11		14		16
Rho^2(0)		0.292		0.299		0.331		0.335		0.345
Rho^2(c)		0.290		0.297		0.329		0.333		0.343
drvcost	-0.300	(-14.1)	-0.405	(-11.0)	-2.010	(-4.4)	-2.350	(-3.3)	-4.45	(-2.9)
fftime	-0.187	(-16.0)	-0.184	(-15.1)	-0.976	(-4.3)	-0.997	(-3.4)	-1.50	(-2.9)
cngtime	-0.299	(-19.9)	-0.296	(-19.1)	-1.530	(-4.5)	-1.48	(-3.4)	-2.25	(-3.0)
rdprice	-0.350	(-16.0)	-0.358	(-15.7)	-2.08	(-4.3)	-2.38	(-3.0)	-3.09	(-2.9)
inpeak	0.612	(2.6)	0.614	(2.6)	1.52	(2.4)	1.67	(2.2)	2.64	(2.0)
offpeak	1.10	(3.4)	1.17	(3.6)	2.63	(2.4)	2.49	(2.2)	3.37	(2.0)
t_malep	-0.674	(-2.5)	-0.705	(-2.6)	-1.86	(-2.5)	-1.93	(-2.2)	-3.28	(-2.1)
asc51	0.188	(2.4)	0.189	(2.4)	0.355	(1.5)	0.329	(1.3)	0.436	(1.5)
costdst			-0.157	(-4.7)	-1.09	(-3.9)	-1.18	(-2.9)	-1.65	(-2.1)
costSP3			-0.346	(-9.4)	-1.96	(-4.0)	-2.38	(-2.9)	-2.59	(-2.8)
ercmp					-1.16	(-4.0)				
cost1_e							-2.00	(-2.7)	-6.21	(-2.7)
fftime_e							-1.12	(-3.1)	-1.74	(-2.7)
cngtime_e							0.904	(2.9)	1.16	(2.7)
cost4_e							-1.48	(-2.3)	-1.96	(-2.7)
cost2_e									7.33	(2.3)
cost3_e									0.924	(1.8)

The variables in the models are defined as:

• *drvcost*; driving cost coefficient in the SP1a (VoT experiment) and SP2a (time of day experiment, TOD).

• *fftime*; free flow time coefficient.

• *cngtime*; congested time coefficient (extra time due to congestion compared to *fftime*).

• *rdprice*; coefficient for road pricing in the road pricing experiment (SP3).

• *inpeak*; dummy variable from TOD experiments (SP2) saying that if the respondent originally travelled in peak hour, then when presented with an out-of-peak alternative he or she might (or might not) prefer to switch. A positive value means that the original time of travel (which is peak) is preferred.

• *offpeak*; dummy variable from TOD experiments (SP2) saying that if the respondent originally travelled in out-of-peak hour, then when presented with a peak alternative he or she might (or might not) prefer to switch. A positive value means that the original time of travel (which is out-of-peak) is preferred.

• *t_malep*; in the TOD experiments (SP2), men who travel originally in the peak are more willing to stay in the peak than women.

• *asc51*; alternative specific constant in the SP3 placed on the left side alternative (i.e. the original route). The positive value means that, when everything else is equal, the respondents prefer the original route.

• *costdst*; cost coefficient calculated via distances in SP1b and SP2b.

• *costSP3*; cost coefficient in the SP3.

• *ercmp*; error component coefficient applied only in model 3. The coefficient was applied on all cost and time coefficients in all SP-experiments. The purpose of model 3 was to discover that the error component improve significantly the model estimations.

• *cost1_e*; cost error component coefficient applied: On all cost coefficients in model 4, and on cost coefficients in the SP1a and SP2a in model 5.

• *fftime_e*; free flow error component coefficient applied on all SP experiments.

• *cngtime_e*; congested time error component coefficient applied on all SP experiments.

• *cost4_e*; road pricing error component cost coefficient in models 4 and 5.

• *cost2_e*; error component cost coefficient in model 5 applied on SP3.

• *cost3_e*; error component cost coefficient (where costs are calculated on distances) in model 5. It is applied on the SP1b and SP2b.

Table 4 shows the VoT's for model 2, which is a MNL model, and table 5 VOT from EC model 5. VoT's in EC models are often calculated as the ratio between the time and cost coefficient (first rows in table 5). However, the distribution of the VoT is more correctly described as the distribution of the time coefficient divided by the distribution of the cost coefficient. The mean of the joint distribution is not the same as the ratio of the means for each distribution. The ratio of two normal distributed terms is Cauchy distributed, for which no estimator for the mean exists. The interpretation of this is that when the denominator (tail of the distribution of the cost coefficient) approach zero, then the VoT's limits infinite. If the distribution contains negative values, then the VoT is negative, which is of course illogical. Due to software limitations, the present study assumed Normal distributed VoT none-the-less.

Table 4. Value of Time (VoT) in DKK/hour in model 2 (MNL model)

Time compo-nent	Cost/time SP	Length/time SP 0.55 Kr/km SP	0,7 Kr/km RP	Road pricing SP
Free flow time	27.3	70.3	89.5	31.9
Congested time	43.9	113	144	51.3

Table 5. Value of Time (VoT) in DKK/hour in model 2 (EC model)

Calculation method	Time component	Cost / time SP	Length / time SP 0.55 Kr/km SP	0,7 Kr/km RP	Road pricing SP
Ratio of coefficients	Free flow time	20.2	54.5	69.4	34.7
	Congested time	30.3	81.8	104	52.1
Simulation	Free flow time	39.4	-	108	44.1
	Congested time	45.5	-	131	49.6

The second row of VoT's in table 5, are simulated with the denominator (cost distribution) truncated at zero. However, as this skewed the cost distribution too much the distribution was truncated symmetrically as well around the mean with the left truncation at zero and the right at twice the mean. However, when the draws at the simulation approach zero, the VoT's still approach infinite. Accordingly a second truncation was done on free flow VoT over 250 DKK and congestion VoT over 400 DKK. The same truncation was decided at the AKTA RP to deal with participants who showed lexicographic behaviour concerning time. The cut-off-values were decided as about twice the maximum VoT of the none-lexicographic participants with the highest VoT. The numerator (time coefficient distribution) is assumed to have a central mean, why truncation or simulation should not be necessary. This was validated by simulation, after which the mean value was used. As seen in table 5, in general the simulated values by this approach were larger compared to the traditional estimation by ratio. It would be expected that the larger EC, the larger difference on the VoT estimates. However, this also increases the likelihood of truncation and the cut-off of maximum VoT values. Hence it is not certain that the VoT would increase (the congestion time in the road pricing SP decrease).

In the SP3, driving costs were presented together with road pricing. This gave a higher VOT (numerically smaller cost-coefficient) than in SP1 and SP2, i.e. the respondents' willingness to pay for time savings increase. However, road pricing had a higher coefficient than marginal cost (both coefficients are still more positive than in experiments without pricing), which indicates that road pricing none-the-less is considered worse than marginal cost.

The congested travel time is weighed more negatively than free flow travel time in all models, as could be expected (which is consistent with prior Danish studies, e.g. Nielsen *et al*, 2002*)*.

The length based experiments showed a much higher VoT than the cost experiment. The results are both showed with the low cost/km. (0.55 DKK/km. - basically a very economic car), and the value used in the RP (0.7 DKK/km. – average car and fuel price as at the time the experiment took place). The interpretation can be, that people in the cost experiment states their thought willingness to pay, while they in real life primarily want to minimise time, and that the length versus time experiment describe this better.

The model also contained a number of dummies. The off-peak dummy revealed a tendency to not change time of day (inertia). This may explain the relatively little change in the time of day of trips in the main AKTA experiment (Nielsen & Jovicic, 2003). It is not surprising that peak hour drivers want to stay in the peak; they have already accepted extra time due to congestion. However, it is a bit surprising that non-peak drivers want to stay out of the peak all-other-things equal. This indicates that they have chosen the non-peak as the best time of day for their specific trip, rather than to avoid congestion in the peak.

The Asc51 variable shows inertia to change route compared to the usual route. T_malep (TOD-game) shows that males are less willing to change TOD than females (opposite the result of Bonsall *et al*, 1998). This is a bit surprising, since woman could be expected to have more constraint in time (shopping, pick up children, etc.).

4.2. Income effect models

Since willingness to pay may depend on income, different parameterised models were tested for this. The more traditional way of modelling income effect is by segmenting the sample by income classes. Several split in classes within the answer intervals between 100, 200, 300, 400, 500, 750 and 1,000 thousand DKK gross house hold income for 2001 were tested.

Data only allowed estimation of two cost coefficients (table 6); for income groups up to DKK 400.000, and above DKK 400.000. 70% of the sample (195 respondents) belonged to the first income group while 30% (84 respondents) belonged to the high income group. The income effect could not be estimated reasonably in the SP1b and SP2b experiments where travel distances were presented to the respondents, i.e. it turned out that lower income group respondents have higher VOT than those with high income.

Table 6. VOT in income dependent EC-models (below / over DKK 400.000), calculated as ratio of coefficients. [is the similar VoT in the models without income dependencies]

	SP1a and SP2a (cost/time trade offs)	SP1b and SP2b (length/time trade offs)	SP3 (road pricing experiment)
Free flow travel time	18.9 / 22.8 [20.2]	Not significant [54.5]	32.1 / 44.8 [34.7]
Congested travel time	28.4 / 34.3 [30.3]	Not significant [81.8]	48.2 / 67.2 [52.1]

The last model type estimated in the study were two models with increasing VoT as function of income, either with an additional $\beta c/I$ term, or $\beta c/I^*c$ (the latter removing the unit problem in the coefficient). However, both formulations had low t-values as well as illogical signs on the coefficients.

4.3. Alternative model formulations

Beyond the model results presented above several other models were tested. The data set is remarkable in the sense that extensive data is available for each respondent since two pure RP and the combined RP/SP questionnaires were available and combinable for all respondents (a unique ID followed all participants through the test). Hereby, several segmentation criterions were possible.

Among the tested models are adding a dummy for peak (fixed and distributed coefficient), combinations of gender and peak (fixed and distributed coefficient), and lognormal distributed cost coefficients (either same distribution or different parameters).

Further a sequence of models where one or more cost coefficients were deflated by income was tested. First linearly increasing coefficients were tested, then dummies for levels in a search for e.g. polynomial relations. At this point, revealed income relations were of illogical relative size or sign; the best model from previous with ‘just’ two income levels of cost coefficients, prevailed.

Finally a model combining the search for non-linear relations while enabling different patterns for different cost components resulted in than the coefficient for road pricing seemed to be increasing by higher income groups. However, the likelihood value for this model (-1642.65) was just slightly improved compared to the MNL model (-1645.71).

5. The RP route choice model

The first step in the model estimation based on the GPS-data was to estimate a route choice model. This model is described in the following.

5.1. Utility function

The route choice model was based on a linear utility function (3), as the SP-model. However, the error term is the sum of error terms at the arcs *a* along the specific route *R*. Since routes can be correlated (sharing links), ε can not be assumed independent distributed among alter-natives and can neither be assumed identical distributed since splitting an arc in two into the digital map would imply a doubled error term. Gamma distributed error terms overcome this, since they are non-negative and overcome the overlapping route problem (Nielsen *et.al,* 2002). The coefficients β could also follow distributions (i.e. error components as in formula 2). However, no distributions were assumed a priori (the empirical distributions were revealed).

5.2. Estimation

$$U_R = V_R + \varepsilon_R = \quad \beta_c RoadPricing + \beta_l Length + \beta_{tff} Time_{ff} + \beta_{tcon} Time_{con} + \sum_{a \in R} \varepsilon_a \qquad (3)$$

The estimation of the route choice model was carried out by running an all-or-nothing route choice model several times for each trip per person. Each run used different combinations of the coefficients in the utility function. The sums of the coefficients were restricted to one, since it is the ratios between the two who determine the choice. This lowered the possible number of combinations, which were pre-defined in a factorial design (Nielsen, 2002) to avoid building the calculation graph dynamically for each run. This speeded up the calculation time in the specific software (ArcGIS).

In the present initial analyses, 48 different combinations of coefficients were run for each trip. And the best fit(s) to the observed route was recorded. The fit was measured as the ratio of the length of the trip which had been fitted to the observed route. Since the network contains 350,000 links, this task was quite calculation demanding. In some (most) cases several combinations gave the same fit (in the extreme case even that a path is both the shortest and fastest). Each fit was then weighted proportional to the number of best fit for the route it tried to match.

It was examined in a pilot study whether an added error term could improve the results (as in formula 3). However, this was seldom the case. The analyses were accordingly conducted by deterministic utility functions for each trip, as this improved calculation time dramatically. The interpretation must be that most of the heterogeneity can be explained by differences in the coefficients in the utility function, and that the remaining unexplained variation can not be explained by a distribution around the deterministic utility function.

The RP-model was based on a marginal cost calculation of fuel (8.5 DKK/l, 12 km. /l) which resulted in a cost per km. of 0.7 DKK, which were higher than the SP (why the results in table 4 and 5 were scaled to the RP-values as well). Using a lower or higher cost/km. scale the VoT linearly, why the same assumptions must be made to compare the results.

5.3. Within person variation

The utility functions for all trips for each car could be compared. About 2/3 of the participants had fairly consistent preferences, whereof the most utility maximising (rational) participant (car) in the experiment got the best route fit for the exactly same combination of coefficients (of the 48 different possible) for all trips. It was in general possible to fit their routes quite well for cars with consistent preferences (in the 70-100% match interval).

However, some participants had a very wide range of preferences. And at the same it was difficult to match the routes (40-60% best match interval). It seems like these participants did not know the network very well. By manual analyses of samples of routes it could be seen that they often followed illogical routes (i.e. non explainable after any criterion the author could figure out). It appears that more work need to be done to explain this group route choices – if it at all is possible.

The best matches for a "random" participant in the experiment are e.g. shown in figure 6, where the preferences vary very much. Using the most likely combination of coefficients (No. 47) may not be a good idea in this case, since it is the weighting of time components only (lexicographic behaviour on the edge of the distribution). Accordingly it is better to use the mean of the values or the median of the distribution. The mean is used in the following.

It was examined whether there was any relationship between the match and socio-economic variables, i.e. if special segments of the population act more rational than others. Such relationships could not be found. There was a week relationship between the fit and VoT though (t stat in regression -1.16), with less fit the higher value of time (decreasing from 70% in average to 60% for those with highest value of time). The interpretation can be that the per-sons with lowest VoT are most cost-aware and thereby examine the options and network more carefully.

5.4. Between person variation

The distribution of the preferences between persons can be revealed by comparing the mean of each person's preferences. Figure 7 and 8 shows the distribution of value of times in the route choice models. The often used assumption in logit models that the coefficients are fixed does clearly not hold. This confirms the SP-based models, where the error component models had much better log.likehood values than MNL. The empirical distributions in the RP-based models are per definition non-negative, as the coefficients have to be positive to make the route choice model work. The distributions are skewed to the right (the mean is larger than the median), and they look a bit logarithmic normal – although the number of observations is too few to determine this for sure. Some of the users have very high value of time, which is due to lexicographic or near lexicographic behaviour (time minimisation only). The observations appear bundled at some few outlier levels, which is because the factorial design of the search algorithm had fewer grid-points in this area of the function.

The VoT from the RP-based model is about 1/3 lower than the SP length experiment based on the same cost per km. The RP VoT are on the other hand are about twice as high than the SP cost experiment. If the empirical distribution in figure 8 is fitted with a Normal distribution, the best fit will most likely be around the "mass" of the distribution and with a right tail fitted to the high values. This again will imply a tail on the left side as well, since the Normal distribution is symmetrical. The VoT calculation by simulation in table 5 truncates the denominator, which to some extent takes care of this, and should make the VoT estimates comparable, i.e. the last two rows in table 5 should be compared with the mean VoT's in figure 7 and 8.

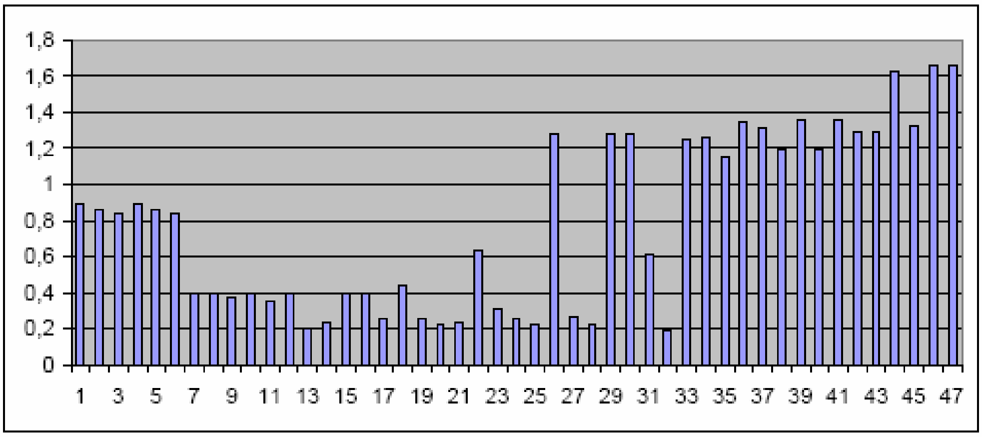

Figure 6 Participant with varying preferences

Another issue is whether the two distributions are independent. The ratio between the two would then contain values below one. As can be seen from figure 9, this is clearly not the case. The smallest ratio is 1.2, i.e. VoT for congestions time is always higher than free flow time. The mean and median is 1.6, which indicates a near symmetrical relationship between the two time-components which was confirmed in a scatter diagram, not shown in the paper. Basically, this illustrate that if one has a high free flow value of time, then the value of time for congestion is even higher. If one has a small free flow value of time, then the congestion value of time is bigger than this. The ratio is in both cases sym-metrically distributed among 1.6.

A regression between the two time-components gave 43.9 in t statistics, i.e. a very strong correlation

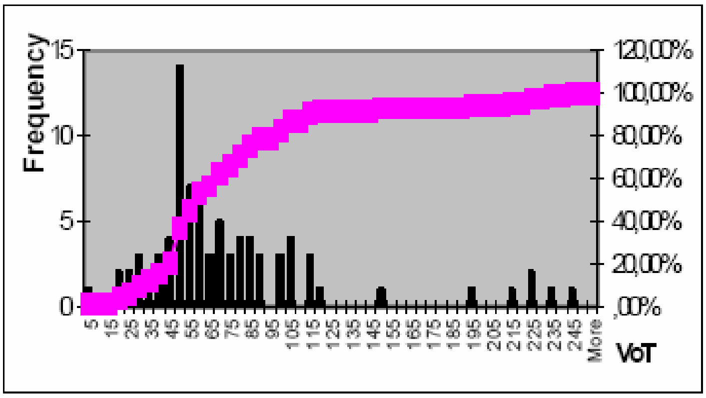

Figure 7 Frequency and accumulated distribution of free flow value of time in DKK. (7.3 DKK equals approximately 1 Euro).Mean VoT = 73 DKK, median = 58.

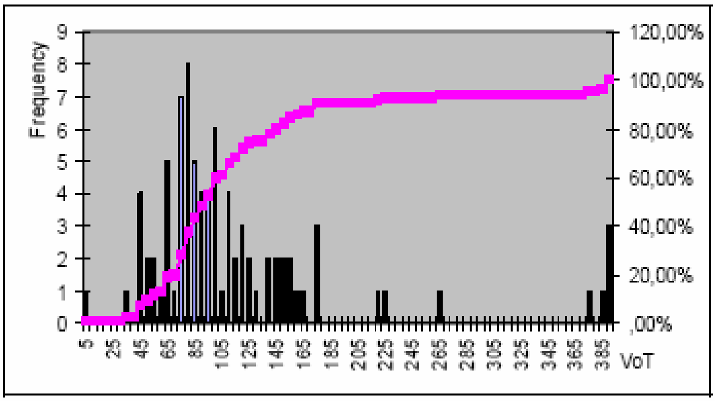

Figure 8 Frequency and accumulated distribution of value of congested time. Mean VoT = 119 DKK, median = 92.

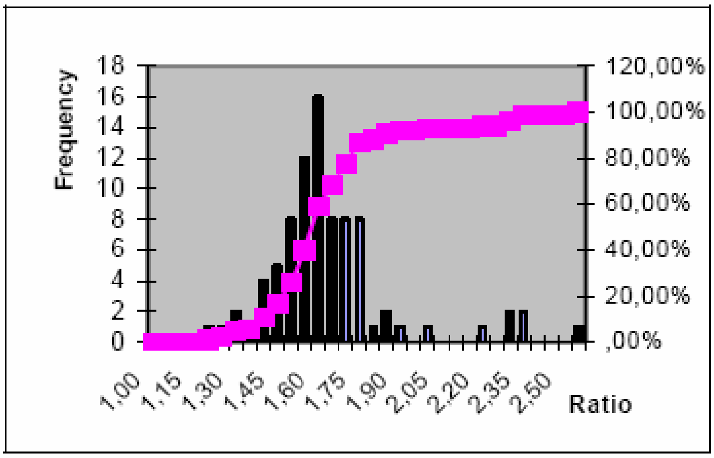

Figure 9 Frequency and accumulated distribution of the ratio between free flow value of time and congested value of time. Mean = 1.6, median = 1.6.

5.5. Estimation of the traffic impact of road pricing

The estimation of route choice under influence of road pricing was done in the same way as the model estimation in the control period. Only the high km.-based scheme has been analysed. It turned out to be difficult to estimate a utility function with one further variable than time due to present software limitations. Another approach was applied instead, where the model was re-estimated assuming the same coefficient on road pricing than marginal cost (length multiplied 0.7 DKK/km). This resulted in a slightly lower VoT than in the control period, i.e. a higher response on road pricing than if assumed equal to marginal cost. This is different from the SP, where the respondents tends to state less change in behaviour than if cost was equal (both the *rdprice* and *costSP3* in the SP were lower, which equals a higher VoT than from the coefficient without road pricing, *drvcost*). The SP model contained in addition a number of dummies for inertia, why it would suppress changes further, while this in the RP is built into the coefficients. It should be noted though that the road pricing experiment in the SP was based on time-of-day decisions (but in a joint model estimation over all 3 experiments), while the RP data analysed all time periods. However, it is surprising that the road pricing in fact lead to such high changes in behaviour in the main RP-experiment.

The opposite was the case in the only comparable experiment known by the authors (Bonsall *et al*, 1998). However, the pricing schemes were defined differently, the urban settings were different, and the pricing level lower. The high km.-based scheme in AKTA varied from 1 DKK per km. to 5 DKK per km. in the peak hours. The marginal driving cost was assumed to be 0.7 DKK (0.55 DKK in the SP). The increase in cost is accordingly quite large. The cost level for public transport is about 1 DRK per km. at the time of the experiment (rough average of the zonal-based price system).

5.6. Income effect

Figure 10 illustrates the VoT as function of house hold income, where the values for each income group have been averaged. The figure shows an increasing relation-ship. However, somewhat surprisingly with the highest VoT level at the third highest income level. This can be due to other socio economic attributes (e.g. higher time budget constraints in families with children, while the average age of the highest income categories must be assumed higher), or higher fixed budget constraints (e.g. housing); or the explanation may simply be small sample size.

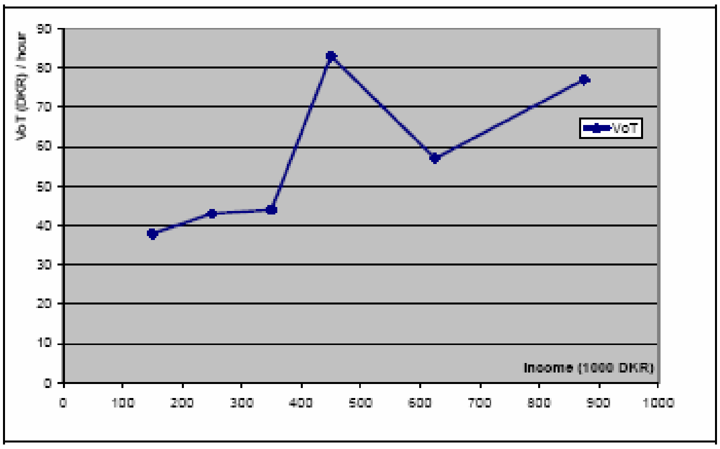

Figure 10. Values of Time as function of house hold income (1,000 DKK).

The figure explain the problems with the SP-model estimation in section 4.2., i.e. where the only possible model had one VoT level under 400,000 DKK and another over, which seems to be the best piecewise specification. However, it doesn't explain why a SP-based model with a linear in-creasing VoT could not be estimated. Except that the SP-data may not include enough observations to estimate this interrelationship.

The relationship between income and VoT has a lot of variation within each income category (figure 11). The regression is quite significant when the line is forced through origin. However, if an intercept is introduced the slope is no longer significant.

5.7. Other explanatory variables

It was possible to estimate some week relationships between the value of time and other socio-economic variables based on the RP-data (opposite on the SP).

There was an increasing VoT with the less restrictions in time, i.e. from 60 DKK with 1) strong restrictions to 80 DKK with 4) no restrictions, over 2) to some extent and 3) to limited extent. The t-stat was 1.74. The relationship is perhaps a bit surprising as one could expect higher

VoT the more time restrictions. However, an explanation can be that low income groups may have more time restrictions (e.g. fix hours of work) or families with small children a tighter money budget. The same tendency was the case for both free flow time and congestion time.

As household size increases VoT decreases (from 100 DKK with one person to 50 DKK with 6 persons, t stat -2.09). The reason here must be tighter budget for children families and maybe also less income, since income increases with age.

The value of time increases with age from 50 DKK for the 20-30 year old group to 95 DKK for the 60-70 years old group (t stat 1.96).

6. Comparison between SP and RP models

It was obvious, that the participants had little feeling with the real marginal or average cost of driving. Dependent on the formulation of the question, different answers were given which were highly inconsistent with each other. The Value of Time (VoT) from the SP-experiments that traded off cost with time differed with about a factor 2-3 from the RP-data. While the SP with trade off of length and times were in the same magnitude as RP, although not equal.

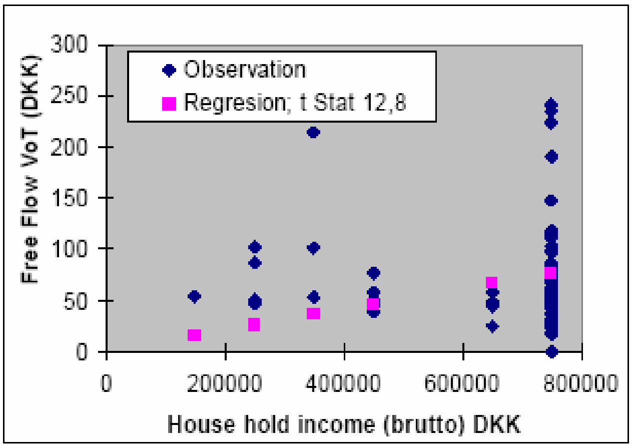

Figure 11 Regression between Value of Time and income.

Using error components improved the SP-based models significantly. A quite large heterogeneity of the coefficients was found (high variance compared to the mean). A similar large heterogeneity was found in the RP-data. The RP-data indicates that the VoT follows logarithmic normal distributions, and that the VoT of different time-components are highly correlated.

About 1/3 of the participants in the RP-experiment showed non-explained behaviour, while the other clearly acted utility maximising. The SP-based logit models assume utility maximisation with an error term. Since the choices here are binary, the unexplained variation will be taken care of by the alternative specific constants and error term. This is more difficult in the RP-based model estimation.

The RP-data showed an increasing VoT with income. This could – with some good will - be assumed linear increasing, or clustered around a level below 400,000 DKK and another level over. The SP-model could only be estimated as a two level VoT due to far less observation than the RP-data set. A split of 400,000 DKK gave the best fit, which was confirmed by the RP-based model.

It is not easy to compare absolute numbers of VoT from the RP and SP-models due to different assumptions on cost/km. as well as problems with calculation of VoT from EC-model with Normal distributions. This will demand a significant software development of the SP-estimation software – which was out of the present projects scope. However, it appears that the SP in the same formulation as the RP (length-time tradeoffs) overestimates VoT slightly. The ratio between different time-components (free flow time and congestion) remains pretty equal, and the magnitude of heterogeneity is similar as well.

The SP states that the value of road pricing is higher than the value of marginal driving cost (fuel manly). However, both cost coefficients became less negative compared to the situation without road pricing (VoT increased). The RP on the other hand showed that the participants changed behaviour slightly more (i.e. a lower VoT) than expected assuming that the value of money were equal. It must be concluded, that respondents tend to underestimate their behavioural changes due to road pricing in the SP-experiment compared to the field experiment, where it is real money. Or they may answer strategically in the SP. This is surprising since they could be expected to change less behaviour in real life due to time-constraints (constraints in their time-budget), habit or lack of reasonable alternatives (as in Bonsall *et al*, 1998).

It was possible to investigate the issues of heterogeneity, income dependent VoT's and other socio-economic variable a bit more in depth using the RP-data than the SP. Although RP-data is more difficult to use for estimation due to the correlation of variables that cannot be controlled as in the SP; it contained far more observations per participant. The SP consist of 3 sub-experiments with 6 trade-offs per participant. While the RP contained between 200 and 1,000 trips per car, each trip with an individually estimated utility function which resulted in about 100,000 observations in the present model.

It must be noticed that the budget for the full RP AKTA experiment was over 10 Mio DKK, while the AKTA SP budget were about 0,5 Mio DKK (including internal funding). The RP contains much more information. But the SP provides – if formulated as length/time experiments and EC-models - good indicators on the behavioural responses.

7. Conclusions

The paper has presented some results and analyses concerning inconsistencies in respondents' answers dependent on how questions are asked, variables defined and whether it is a RP or SP experiment. All questions were asked to the same group of participants, i.e. the same person faced different questions within the same SP, followed by the RP over a minimum of 2x8 week period. As the RP data was collected by GPS technology, the behaviour could be measured very accurately.

The respondents evaluation of driving cost were addressed by various approaches. They were asked to confirm or modify cost and length estimates. Although cost a priori was calculated as a function of the length, only 18% of the participants modified this while 43% modified the length. They were then asked their average driving cost per km. This turned out to be highly inconsistent with the accepted cost/length from before. Finally, they were asked which elements they had included in the calculation of cost/km. This again was for many respondents highly inconsistent with their prior answers. It appears accordingly, that car users have very vague and erroneous knowledge on the cost of driving, while they have a fairly good knowledge on the travel length and time.

The SP-experiment then focused on Value of Times (VoT), where half of the respondent's chosen between time and length and the others between times and cost. Cost was defined proportional to the length and estimated based on the usual trip. The resulting VOT's differed a lot by the two methods with up to a factor 3. However, the ratio between free flow VoT and congestion VoT was comparable. The SP-model based on length/time trade-off and cost-estimate based on true marginal driving cost gave almost identical VoT as the AKTA RP-data.

The SP also investigated road pricing on an existing trip, which were then charged. This lead to higher VoT than the RP route choice experiment (about 50%), while the coefficient on road pricing was about 20% higher than the pure cost coefficient. The interpretation must be that people dislike road pricing more than other out of pocket costs but that their behaviour none-the less changed to a less degree than expected. This must be interpreted as some inertia for changes (which was also described by a dummy in the SP-model), which however is somewhat counterintuitive compared to the higher value of money (coefficient) on road pricing.

The AKTA RP provides a clearer conclusion; the participants changed to less degree (about 20%) behaviour than expected if the value (coefficient) on road pricing where the same as marginal driving cost. The estimated utility functions confirmed this; both their coefficients, and value of times. The interpretation must be that respondents in SP may overestimate their changes in behaviour – or answer strategically – but that they faced with real life restrictions in time and space none-the-less stay with their existing pattern (maybe due to habit, convenience, or lack of knowledge on alternative options).

All models were initially estimated as traditional logit models, followed by a re-estimation in a mixed-logit framework. The ML-models based on SP were estimated by maximum simulated likelihood. Different distributions were applied – first normal in different combinations on the coefficients, then logarithmic normal, and finally simultaneous distributions (to allow for correlation between the coefficients). The different assumptions did alter the results (values - not signs or magnitudes).

The RP-data contained enough information to estimate a model per participant. The empirical distributions of the coefficients within and between respondents could accordingly be derived. And statistical distributions were fitted to this, where logarithmic normal gave the best result. In this context it is interesting to note, that some of the car-users showed lexicographic behaviour – i.e. that the cost coefficients were zero within their available choice set and the VoT in principle hereby were infinite. The distribution was accordingly skewed a bit more to the right than the Logarithmic normal.

Another conclusion from the RP-data was that about 2/3 of the car-users had a clear utility maximising behaviour, i.e. that their choices could be explained with a good fit by a utility functions where the coefficients between trips had a reasonable small variance. However, the remaining 1/3 of the car-uses had a very high variance on the coefficients - they sometimes minimised costs and other times time - at the same time as a less good fit was obtained to the observed routes.

Both the RP and SP models were tested for income effect. Both as dummies, piecewise coefficients, and a full specification of income effect in the indirect utility function following a micro economic framework (cost-dependent cost-coefficients). By fiddling with the interval, the SP-model could be estimated with VoT in two intervals (all other models gave illogical signs and small t-values). The RP-data could reveal VoT's for each of the 9 income-intervals that people have provided information for. The most reasonable functional form to explain this did reproduce the findings from the SP.

The overall conclusion is that car users are much better at estimating length than cost. There are a lot of heterogeneities in behaviour and preferences, which best can be fitted by logarithmic normal distributed coefficients. These are most likely very correlated between different time-components. If designed the right way SP can lead to results in the same magnitude as RP.

However, in the worst cases there could be a factor 2-3 in difference between the results and even signs of impacts could differ.

Acknowledgement

Mai-Britt Herslund is thanked for the work on SP-design and the interviews, Christian Würtz on work collecting and processing the GPS-data, Paolo Menegazzo on work on the route choice model estimator and data processing, and Goran Jovicic are thanked for work on the SP and the jointly authored paper Nielsen & Jovicic (2003).

The AKTA project is primarily financed by EU's 5th Framework Programme and the municipality of Copenhagen, with co-funding to the research part from the Technical University of Denmark (DTU). Principle contractor of AKTA is the City of Copenhagen, from which Poul Sulkjær is thanked. DTU, The Danish Road Directorate and The Ministry of Transport are assistant contractors. The AKTA SP was funded by the Danish Transport Council, and the remaining work on this was allowed to be finalised by self-funding from DTU and the Danish Transport Research Institutes.

References

Bonsall, P., Cho, H-J, Palmer, I and Thorpe, N. 1998. Experiments to determine differences in drivers' response to a variety of road user charging regimes. PTRC. September 1998.

Nielsen, Otto Anker; Frederiksen, Rasmus Dyhr, & Daly, 2002. Andrew. A stochastic multi-class road assignment model with distributed time and cost coefficients. Networks and spatial economics. 2 327-346. Kluwer.

Nielsen, O.A,. and Jovicic, G., 2003. The AKTA road pricing experiment in Copenhagen. 10th International Conference on Travel Behaviour Research. Proceedings, session 3.2 Valuation/Pricing. Lucerne, Switzerland, August.