USES OF THE LOGIT SCALING APPROACH IN STATED PREFERENCE ANALYSIS

Mark BRADLEY
Associate
Hague Consulting Group
The Hague - The Netherlands

Andrew DALY
Director
Hague Consulting Group
The Hague - The Netherlands

INTRODUCTION

The scaling approach is a statistical estimation method which allows for differences in the amount of unexplained variation in different types of data which can then be used together in analysis. In a number of recent papers, this approach has been tested and recommended in the context of combining Stated Preference and Revealed Preference data. Section 1 provides an overview of the approach and the recent literature.

The scaling approach may be useful in a wider range of contexts than has been discussed hitherto. For example, there may be systematic differences in the variance of choices within a single Stated Preference data set due to the way in which the hypothetical choice situations are presented or the responses are obtained. Sections 2 and 3 report the results of two such case studies - looking at rank order effect and fatigue effect, respectively.

A brief summary and conclusions are given in Section 4.

1. THE SCALING APPROACH

This section introduces the scaling approach: first in terms of statistical theory and then in terms of recent experience.

1.1. Statistical background

The objective of discrete choice statistical estimation techniques such as logit and probit is to estimate a utility function which best predicts the choices in an observed sample. Supposing that we have two types of data, the utility functions to be estimated may appear as:

$$u_1 = \beta.X_1 + \alpha.Y + \varepsilon_1$$

$$u_2 = \beta.X_2 + \gamma.Z + \varepsilon_2$$

where:

β is a vector of parameters to be estimated, assumed to have the same values in both data sets;

X_1, X_2 are vectors of observed values of variables common to the two data sets;

Y, Z are vectors of observed variables which may be specific to one data set or the other;

α, γ are vectors of parameters to be estimated for the data-specific independent variables;

$\varepsilon_1, \varepsilon_2$ represent the amount of residual, unexplained variance of the utility functions in the two data sets.

The correct values of β are assumed to be identical for all choice observations used to estimate the utility functions. In practice, this assumption is not too restrictive, as the data-specific variables in Y and Z can be used for effects which apply only to certain types of observations.

The random error terms, ε_1 and ε_2 are assumed to incorporate all unobserved or unspecified effects on the choices. These terms are assumed to be (approximately) normally distributed with a mean of 0 and a variance which is identical across all relevant observations. If, in reality, certain types of observations systematically have more or less unexplained variance than other types of observations, this can lead to two main types of problems:

(1) The model parameters may be biased: Observations with systematically larger random variance should receive less "weight" in the estimation. If this is not the case, the estimated β coefficients for the observed effects may be biased.

(2) The model elasticities may be too high or too low: In application, the sensitivity of the model predictions to changes in the X variables is a function of the overall scale of the β parameters relative to the variance of error term ε. If, therefore, observations are used in estimation which have a variance much different from that in the context for which predictions are made, the model will tend to over- or underpredict actual changes in choice probabilities.

In the notation above, such problems could occur if we attempt to estimate the utility functions u_1 and u_2 in a single model which assumes constant variance across observations, when in reality ε_1 and ε_2 are not the same. The scaling approach addresses this problem by allowing different types of observations to have different error variances within a single model. Suppose that

$$\theta^2 = \mathrm{var}(\varepsilon_1) / \mathrm{var}(\varepsilon_2)$$

Then scaling the utility of data set 2, $u_2' = \theta . u_2$, will allow it to be estimated in a single model with u_1 in an efficient and unbiased manner. This notation can be extended to apply to more than two types of data.

The two main challenges to applying this approach are (1) determining how the different utility functions, u_1, u_2, etc. should be specified to separate different types of observations and the different types of variables that might influence them, and (2) incorporating the method within a feasible and accessible model estimation procedure that can estimate the scale factor(s) θ at the same time as the other unknown parameters (β,α,γ). A more detailed treatment of the method can be found in much of the recent literature which is reviewed below.

1.2. Relevant literature

Discrete choice analysis has for many years been the dominant estimation approach used in disaggregate travel demand modelling. A comprehensive treatment of this approach is given by Ben-Akiva and Lerman (1985). The examples are mainly oriented toward analysis of Revealed Preference (RP) data based on observations of actual travel choices.

During the 1980's, increasing reliance was placed on Stated Preference (SP) data based on stated choices under hypothetical travel contexts. Due to the use of controlled, experimental survey contexts, such data have proven very effective in estimating the relative parameters (β's) for travel demand models. Direct application of such models in forecasting actual choices may not be appropriate, however, because of the scale sensitivity problem mentioned above.

A number of authors have discussed the reasons why one might expect differing amounts of unexplained variance in choices from RP and SP survey contexts (Bates 1988; Morikawa, et.al. 1990; Bradley and Kroes 1990a, 1990b; Daly and Ortúzar 1990). In the sense that SP experiments can be to some extent isolated from uncontrolled variables, one usually expects less residual variance in SP data, and thus a higher scale on the coefficients. To make SP models more applicable to actual choices (and more relevant in general), there has been a growing emphasis on basing SP choice contexts as closely as possible around actual choice situations recently faced by the respondents.

With the objective of combining the best aspects of both RP and SP data, it was proposed that both types of data be used together in analysis. If the differences in variance in the data were not accounted for, however, such analysis would encounter scale problems of the type we have mentioned.

Ben-Akiva and Morikawa (1990) developed a procedure for efficiently estimating the scale differences between different data sources. This involved estimating a model using a joint likelihood function for observations from two or more data sources. As long as the utility function for each data source has at least two β parameters in common with those for the other data sources, a relative scale factor θ can be estimated for each utility function. This approach was programmed in the GAUSS language as a binary probit estimation procedure and applied to a data set from the Netherlands which contained self-reported RP observations as well as observations from two separate SP experiments. The results are reported in Morikawa (1989).

The multinomial logit scaling approach was introduced in a paper by Bradley and Daly (1991). This approach uses an existing "tree" logit estimation package (ALOGIT) to carry out the one-step estimation procedure (Daly 1987). The paper by Bradley and Daly presents two case studies. The first used the same Dutch data set as used by Morikawa and gave almost identical results. The second was based on much larger RP and SP data sets from Australia. The second case was a somewhat more rigorous test, as the RP and SP data sets were collected from different samples and survey instruments.

In addition to its speed and accessibility, an advantage of the logit-based scaling approach is the ability to estimate multinomial and nested (tree) models with large numbers of alternatives. This capability was necessary for the Australian case study which involved a model of choice from among five long-distance travel modes. In later analyses, not yet reported, nested models were used with all passenger modes in a single nest versus private car.

The scaling approach may be useful in other contexts besides the mixed analysis of RP and SP data. Ben-Akiva and Morikawa (1990) note other possible contexts, such as the mixture of data from different types of RP data - e.g. household surveys, roadside interviews, traffic counts, etc.

The internal sources of variance in SP data can also be investigated using the scaling approach. One of the most often-discussed features of SP data is the repeated measures aspect: the complex mixture of within-person and between-person variance in the responses. Unfortunately, this aspect is not easily addressed using the scaling approach because there is often no way of separating the types of observations - each SP choice is subject to both within- and between-person variation. The extreme option of treating each person in the sample as a separate data set with its own scale factor is not likely to be practical or useful. A promising approach for dealing with some aspects of the repeated measures issue may be the random effects models sometimes used in analysing longitudinal data (e.g. Meurs 1991).

Certain aspects of the repeated measures problem can be addressed, however, if we can account for that part of the within-person variance which arises from the survey instrument itself. One could imagine, for example, that the survey method could introduce differences in variance between observations within a single choice experiment. For a ranking exercise, people may pay more attention (less "random" error) to the ordering of the best options than they do to the ordering of the worst options. For a series of pairwise choice exercises, respondent "fatigue" may cause people to make choices less carefully as the number of exercises increases.

The following two case studies apply the scaling approach to investigate the rank order and fatigue effects just hypothesised. The questions which we address are: (1) do such effects seem to exist and (2) if so, what effect do they have on the estimation results?

2. CASE STUDY 1: RANK ORDER EFFECTS

The first case study is based on SP data collected during a study conducted for Stockholm Transport, described in Widlert, et.al. (1989). The data was collected during home interviews with over 300 people who commute by bus. Separate SP experiments included variables related to bus service levels, bus stop facilities and bus vehicle factors respectively. These experiments were administered to each person in random order during the interviews.

Each experiment contained nine alternatives, created using standard fractional factorial designs with Hague Consulting Group's SPEED software. Two additional alternatives were created - one with the best possible levels of all variables, and the other with the worst possible levels. These two extreme cards were placed at the ends of a metre stick. The respondent then ranked the other nine design option cards in preference order by arranging them between the best and worst options.

Analysis was done using the exploded logit method in which a ranking of N alternatives is treated as N-1 independent observations: 1st chosen over 2nd through Nth, 2nd chosen over 3rd through Nth, ..., (N-1)th chosen over Nth. This analysis method has often been subject to criticism; although the required distributional assumptions for logit estimation are satisfied, it is questionable whether the utility assumptions underlying discrete choice methods such as logit are applicable to rankings (Daly 1990).

Estimation results for the three experiments are presented in Tables 1 to 3. The first ("Basic") model in each table was estimated using only the design variables in dummy (0/1) variable form. All three experiments had 3 levels of fare, allowing estimation of two fare variables. Each experiment also contained three other variables, each with 2 or 3 levels. Each experiment was somewhat different in terms of the number of levels and the experimental design used. In this paper, we do not have space to discuss the individual variables themselves in any detail.

In the Basic models, all coefficients are significant with the expected signs, except for "automated ticketing only" in Table 2, which gave a somewhat ambiguous result, with some people for it and some against it. Note that the significance (t-statistics) for the Basic models may be overstated because the repeated measures nature of the data has not been accounted for in any way.

For the second ("Scaled") model in each table, the data was essentially split into eight separate data sets, with one corresponding to each "chosen" rank in the order - all 1st choices as chosen, all 2nd choices as chosen, etc. The first rank was set as the "base" data set, and the logit scaling technique was used to estimate seven scale factors for ranks 2 to 8 relative to rank 1. These scale factors are shown in the tables along with the t-statistics relative to 1.0 (no difference in variance). All scale factors are significantly less than 1.0, indicating that the amount of unexplained variation increases with lower rankings. For all three models, the scale factors for ranks 2 to 4 are in the range near .50 to .75, while the scales for ranks 5 to 8 are in the range near .20 to .50. The decreasing trend can be seen clearly in Figure 1.

Table 1: Experiment 1 - Service Levels
302 respondents, 2416 observations

Model	Basic		Scaled		Simulated	
Log-Likelihood(b)	-3161.2		-3073.2		-3144.0 (vs. -3145.9)	
Coefficients (t-statistics w.r.t. 0)						
Fare +20%	-.952	(15.5)	-1.425	(8.9)	-.831	(9.6)
Fare -20%	.550	(9.7)	1.241	(9.7)	.629	(9.0)
More punctual	.792	(14.0)	1.468	(9.8)	.768	(10.0)
Less punctual	-.939	(15.3)	-1.720	(10.7)	-.948	(9.7)
One interchange	-.561	(11.1)	-1.523	(8.2)	-.554	(7.6)
Travel time -20%	.531	(10.5)	.656	(6.7)	.578	(9.0)
Scale factors (t-statistics w.r.t. 1)						
Rank 1			1.000	(-.-)	1.000	(-.-)
Rank 2			.725	(3.3)	1.074	(0.7)
Rank 3			.704	(3.7)	1.157	(1.2)
Rank 4			.792	(2.1)	1.098	(0.8)
Rank 5			.472	(8.5)	1.009	(0.1)
Rank 6			.217	(19.1)	1.003	(0.0)
Rank 7			.431	(10.0)	.939	(0.5)
Rank 8			.241	(12.6)	1.032	(0.2)

Table 2: Experiment 2 - Bus Stop Facilities
294 respondents, 2352 observations

Model	Basic		Scaled		Simulated	
Log-Likelihood(b)	-2915.7		-2846.4		-2880.9 (vs. -2883.6)	
Coefficients (t-statistics w.r.t. 0)						
Fare +10%	-.902	(14.9)	-1.496	(10.7)	-.843	(12.2)
Fare +20%	-1.861	(26.6)	-3.436	(12.3)	-1.740	(15.4)
Bus shelter	1.840	(29.2)	3.802	(11.1)	1.779	(14.1)
Real time info.	1.131	(19.8)	1.693	(11.5)	1.034	(14.4)
Automated ticket	-.242	(4.6)	-.054	(0.4)	-.210	(4.1)
Scale factors (t-statistics w.r.t. 1)						
Rank 1			1.000	(-.-)	1.000	(-.-)
Rank 2			.631	(5.8)	1.077	(0.9)
Rank 3			.391	(13.4)	1.125	(1.3)
Rank 4			.493	(9.4)	1.043	(0.4)
Rank 5			.533	(7.4)	1.053	(0.5)
Rank 6			.317	(12.1)	1.269	(1.6)
Rank 7			.276	(11.8)	.981	(0.2)
Rank 8			.239	(11.1)	1.006	(0.0)

Table 3: Experiment 3 - Vehicle Factors
300 respondents, 2400 observations

Model	Basic		Scaled		Simulated	
Log-Likelihood(b)	-3021.6		-2904.4		-2995.0 (vs. -3001.1)	
Coefficients (t-statistics w.r.t. 0)						
Fare +20%	-.584	(10.1)	-2.150	(7.6)	-.528	(7.9)
Fare +40%	-1.871	(26.4)	-5.574	(9.0)	-1.856	(11.9)
Stand 2 minutes	-1.063	(18.0)	-3.201	(8.9)	-1.047	(11.8)
Stand 10 minutes	-1.974	(28.5)	-5.748	(9.7)	-1.995	(12.4)
Clean inside	.374	(6.4)	1.650	(5.8)	.343	(5.0)
Clean inside+out	.387	(6.5)	2.915	(6.0)	.344	(4.2)
Destination signs	.270	(5.4)	.100	(0.6)	.224	(4.4)
Scale factors (t-statistics w.r.t. 1)						
Rank 1			1.000	(-.-)	1.000	(-.-)
Rank 2			.541	(7.8)	1.043	(0.4)
Rank 3			.430	(11.0)	.931	(0.7)
Rank 4			.457	(8.7)	.920	(0.8)
Rank 5			.281	(17.1)	1.141	(1.1)
Rank 6			.125	(35.1)	1.074	(0.5)
Rank 7			.190	(26.9)	1.329	(1.8)
Rank 8			.190	(27.9)	.912	(0.5)

In terms of model fit, the addition of seven scale parameters adds 88, 69 and 117 log-likelihood units for the three experiments with respect to the Basic model. These are highly significant improvements according to the likelihood ratio test.

The t-statistics of the design variables are generally reduced by one half to one third relative to the Basic models. One could consider these values to be closer to the "true" significance of the effects, since more aspects of behaviour have been accounted for. Note that the two smallest coefficients, for automated ticketing and destination signing, are no longer significant.

Not only does the scale and significance of the coefficients change in the Scaled models, but the relative magnitudes sometimes shift as well. In Table 1, for instance, interchange becomes more important relative to the other variables, while travel time becomes less important. In Table 3, cleanliness of the buses becomes more important relative to the other variables, while destination signing becomes less important.

Grouping the data from different rankings thus appears to have biased some of the results away from the values obtained when scale differences are accounted for. An even stronger test is, instead of scaling, to estimate completely separate models for each of the eight ranks and look at the change in total likelihood. Such tests (not shown in the tables) gave improvements of 175, 108 and 180 log-likelihood units with respect to the three Scaled models, for the addition of 36, 30 and 42 parameters, respectively. These are significant improvements, indicating that the scale difference alone does not explain all of the differences between the ranks. Ben-Akiva, et al. (1991) also found significant parameter differences corresponding to rank order.

The results above certainly have implications for the use of rank-order SP data and exploded logit. These implications are discussed in section 4. First, however, we were concerned that the results might be an artefact of the analysis method used. To test this, the estimated utility functions from the Basic models, including random error component, were used to simulate rank-order responses which were substituted for the actual responses in the SP data sets. The simulated responses assumed no rank-order scale effect: the same error variance was assumed for every alternative (simulated error terms were drawn randomly from a standard Gumbell distribution).

The same Basic and Scaled model specifications used in Tables 1 to 3 were estimated using the simulated rankings for all three experiments. The Basic model results (not shown) were essentially the same as for the actual rankings in terms of coefficients, t-statistics and likelihoods.

The Scaled Simulated model results are shown in the final column of each table. The coefficients remain close to those used to simulate the data, and, contrary to the models on the actual data, the rank-specific scale factors are in no cases statistically different from 1.0. Figure 1 shows the contrast between the actual and simulated scale results more clearly. Furthermore, the improvements in log-likelihood over the Basic models are only 2, 3 and 6 units, not significant for the addition of 7 scale parameters. Further tests estimating completely separate parameters for each simulated rank (not shown) gave improvements of 15, 11 and 22 log-likelihood units over the Scaled models - again not significant. These results provide strong evidence that the rank-order effects are due to differences in the way in which real respondents make decisions at different points in the ranking process.

3. CASE STUDY 2: FATIGUE EFFECTS

The second case study is based on SP data collected for the Dutch Railways to study train/car mode choice for intercity travel in the Netherlands, described in Bradley et al. (1988). This same data has been used in the context of joint RP-SP analysis by Morikawa (1989) and Bradley and Daly (1991). Here, we use only the data from the first SP experiment which offered choices between alternative train service options.

The survey was administered using Hague Consulting Group's MINT interview software. MINT used the four design variables to create a fractional factorial design of nine alternatives. Varying orthogonal sets of options were selected for different respondents on a random basis. Respondents were presented with a series of pairwise choices from among these nine options. The first pair offered was a "dominant" choice, where one option was clearly superior to the other. This first pair served as a lead-in to the experiment where the interviewer could check whether the respondent understood the choice task. No other such dominant pairs of options were shown to respondents. From the second choice onwards, pairs of alternatives were presented in random order, until the point where MINT could infer a full preference ordering across the alternatives. On average, each respondent completed about 13 pairwise choices, with nearly all respondents completing between 10 and 16 choices.

Figure 1
Scale Factors for Exploded Rank Models

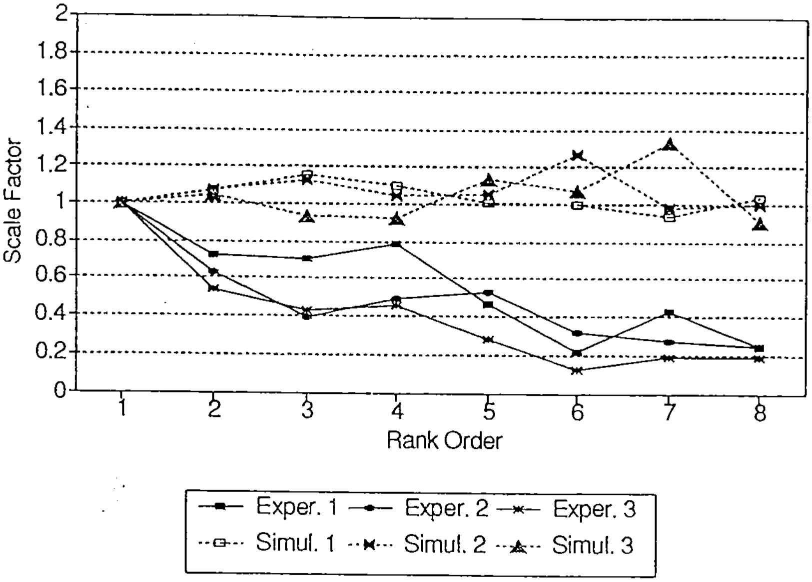

Figure 2
Scale Factors for Fatigue Effect Models

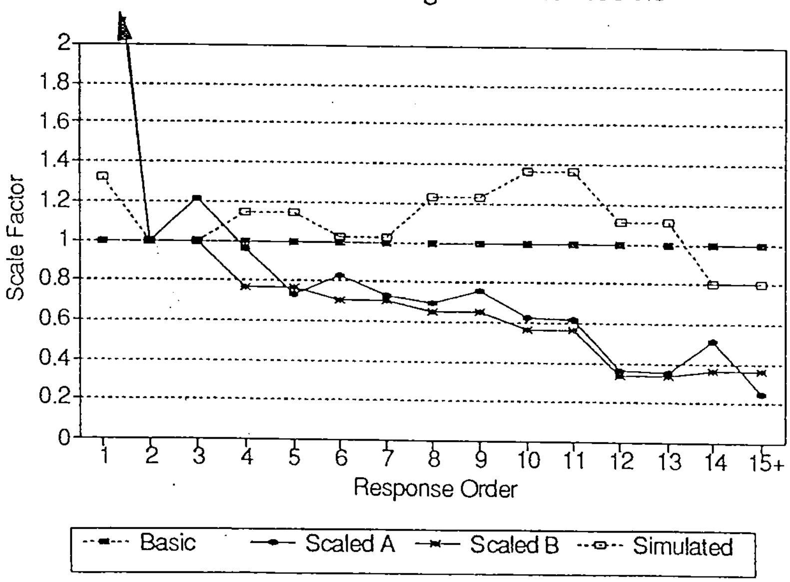

Table 4: Train Service Pairwise Choices
234 respondents, 2929 observations

Model	1- Basic	2- Scaled A	3- Scaled B
Log-Likelihood	-1724.2	-1668.2	-1670.1
Coefficients (t-statistics w.r.t. 0)			
Fare (fl)	-.1484(19.9)	-.1723 (6.3)	-.1900 (9.2)
Time (min)	-.0287(10.7)	-.0332 (5.4)	-.0356 (6.6)
Transfers (N)	-.3263 (5.5)	-.3412 (3.3)	-.3884 (3.9)
Comfort Level	.9457(14.6)	1.1400 (5.8)	1.2500 (8.1)
Scale factors (t-statistics w.r.t. 1)			
Response 1		3.392 (2.8)	3.096 (3.2)
Response 2		1.000 (-.-)	1.000 (-.-)
Response 3		1.219 (0.8)	" "
Response 4		.962 (0.2)	.764 (1.9)
Response 5		.735 (1.5)	" "
Response 6		.830 (0.9)	.709 (2.4)
Response 7		.729 (1.5)	" "
Response 8		.694 (1.7)	.657 (2.9)
Response 9		.752 (1.3)	" "
Response 10		.629 (2.0)	.565 (3.6)
Response 11		.616 (2.3)	" "
Response 12		.365 (3.8)	.332 (5.7)
Response 13		.355 (3.2)	" "
Response 14		.588 (1.7)	.363 (4.7)
Response 15+		.242 (4.0)	" "

Estimation results are presented in Table 4. Model 1 is the "Basic" model, using binary logit with linear functions of the four design variables, each of which had three customised levels in the experimental design. The coefficients are all significant with the expected sign.

The second column shows the "Scaled A" model. Here, the data was separated into 15 groups: all choices which were done 1st, all which were done 2nd, and so on. Because the first choice faced by each respondent was an obvious "warm-up" comparison, the second choice was specified as the base data type. Fourteen scale factors were estimated for the other response orders, relative to the second response. The same four design variable coefficients were specified to apply to all responses.

The resulting improvement in log-likelihood relative to the Basic model is 56 likelihood units, which is highly significant for the addition of 14 parameters. The scale for the 1st response is quite high, as one would expect - there is little chance for error to affect such obvious dominant choices. The scale for the 3rd response is somewhat higher than for the second, but from the 4th response onward the scale is consistently less than 1.0. Although the standard error for the scale factor estimates are fairly high due to the limited sample sizes, one can see in Figure 2 that a clear trend is present. A "fatigue" effect (higher unexplained variance) appears to set in around the 5th response and to become much stronger around the 12th response. As the scale of the first "warm-up" choice appear to be quite different than the rest, the models were reestimated omitting the first observation per respondent. The results (not shown here) did not change noticeably.

In the third model in Table 4 ("Scaled B") the response orders are grouped into pairs, so that only 7 scale factors are estimated. Compared to the "Scaled A" model, the loss in log-likelihood is only 2 units with 7 fewer parameters. The scale parameters are also now significantly different from 1. Although this model gives lower standard errors of the parameters than "Scaled A", the trend in scale factors in Figure 2 appears virtually unchanged.

In contrast to the rank-order effect results presented earlier, the scaling approach has almost no effect on the relative magnitude of the coefficients. The inferred values of time for the three models in Table 4 are 11.60, 11.56, and 11.24 guilders per hour. The t-statistics for the Scaled B model are one half to one third lower than for the Base model - a result similar to that obtained in the ranking case study.

Also in contrast to the rank-order results, a model which, instead of scaling allowed all four design coefficients to be estimated separately for each response order showed no significant improvement over the scaled models (33 log-likelihood units for the addition of 45 parameters - results not shown). Possible reasons for the contrasts between the rank order and fatigue results are discussed in section 4.

As before, we wanted to be certain the results in Table 4 are not an artefact of the scaling approach itself. The estimated utility function from the Basic model, assuming a random error component with constant variance, was used to simulate pairwise choice responses which were substituted for the actual responses in the SP data. The Basic and Scaled B model specifications in Table 4 were then estimated using the simulated choices. Although the results are not shown, none of the estimated scale factors were significantly different from 1.0. Figure 2 shows the clear difference between the results for the actual and simulated choices. The scaling approach on the simulated data increases model fit by only 4 log-likelihood units, which is not significant for the addition of 7 parameters. Again, the simulation tests provide evidence that the "fatigue" effect is a real-world phenomenon.

4. SUMMARY AND CONCLUSIONS

The scaling approach is an efficient estimation approach to account for differences in the amount of unexplained variance when using different types of data together in estimation. The logit scaling approach has been presented and tested in recent literature in the context of combining Revealed Preference and Stated Preference data. In this paper, we have investigated the use of the logit scaling approach to account for survey effects within Stated Preference data from a single experiment. In two case studies, we have tested for the presence of a rank-order effect in rankings data and of a "fatigue" effect in repeated pairwise choice data.

The results from the case studies show a number of similarities. Scale effects appear to exist in both cases: the amount of unexplained variance is shown to increase as rankings become lower, and as the number of pairwise choices completed becomes greater. In both cases, the t-statistics of the design variable coefficients generally decrease by one half to one third compared to the basic "naive" estimation results. The overall fit of the models, however, substantially improves due to the addition of the scale parameters. In both cases, these effects could not be reproduced using simulated response data; indicating that the rank-order and fatigue effects are caused by the influence of the experimental tasks on the respondents. Although it may be too soon to make general conclusions about the existence of such effects based on these case studies alone, we can add that the types of experimental designs and interview methods used to collect the data are very typical of those used in many recent transport SP studies.

We also obtained some contrasting results from the two case studies. For the pairwise choice data, the addition of the scale factors to account for respondent fatigue did not significantly change the relative magnitude of the model coefficients. For the rank-order data, this was not the case: some coefficients became more or less important relative to the others, and estimating separate models for different positions in the rank order gave a significant improvement in likelihood relative to using only scale factors. Similar results were obtained by Ben-Akiva et al. (1991).

The pairwise choice data was collected using a computer-based approach which used some presented the pairs of alternatives in a different random order to each respondent. As a result, any adverse influence of the order-related fatigue effect may have been randomised out of the data. If a survey approach had been used where everyone received the same pairs of alternatives in the same order, the findings may have been different. For ranking exercises, where the ordering is determined by the respondent, it would be difficult to completely eliminate any relationship between the rank order and the levels of the design variables. Some of the effects may be eliminated, however, by using block experimental designs which present different groups of respondents with different sets of alternatives for ranking.

With regard to pairwise choice data, the results indicate that strong fatigue effects should be avoided by not offering more than 10 or so choice comparisons within a single experiment. This number, of course, may vary with the difficulty of the choices offered and the total length of the survey. If one can randomise the order in which pairs of alternatives are presented, it is probably not essential to use the scaling approach, as the relative magnitudes of the model coefficients will not be biased. This sort of randomisation is greatly facilitated by computer-based interviewing.

With regard to ranking, the results suggest that one should not go beyond using the first 3 or 4 ranks as choices in exploded logit, that one should check the extent to which the results change as more ranks are used, and that the scaling approach should be used to avoid biases. If possible, different blocks of alternatives should be given to different groups of respondents for ranking, and the ranking task should be administered in a way which encourages respondents to give equal attention to the ranking of the more-preferred and less-preferred alternatives.

BIBLIOGRAPHY

Bates, J. (1988) "Econometric Issues in SP Analysis". Journal of Transport Economics and Policy. V 12(1). pp. 59-70. Bath.

Ben-Akiva, M. and Lerman, S. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press. Cambridge, Mass.

Ben-Akiva, M. and Morikawa, T. (1990). "Estimation of Travel Demand Models from Multiple Data Sources". 11th International Symposium on Transportation and Traffic Theory, Yokohama.

Ben-Akiva, M., Morikawa, T. and Shiroishi, F. (1991). "Analysis of the Reliability of Preference Ranking Data". Journal of Business Research. V. 23. pp. 253-286.

Bradley, M. and Daly, A. (1991). "Estimation of Logit Choice Models using Mixed Stated Preference and Revealed Preference Information". 6th International Conference on Travel Behaviour. Quebec.

Bradley, M.A., Grosvenor, T. and Bouma, A. (1988). "An Application of Computer-Based Stated Preference to Study Mode Switching in the Netherlands". PTRC 16th Annual Summer Meeting- Proceedings of Seminar D. University of Bath.

Bradley, M.A. and Kroes, E.P. (1990a) "Forecasting Issues in Stated Preference Survey Research". 3rd International Conference on Survey Methods in Transportation. Washington.

Bradley, M.A. and Kroes, E.P. (1990b). "Simultaneous Analysis of Stated Preference and Revealed Preference Information". PTRC 18th Annual Summer Meeting- Proceedings of Seminar H. University of Sussex.

Daly, A. J. (1987). "Estimation of 'Tree' Logit Models", Transportation Research, Vol. 21B, pp 251-267.

Daly, A.J. (1990). "Integration of Revealed Preference and Stated Preference Data in Model Estimation". Banff Invitational Symposium on Consumer Decision Making and Choice Behavior. Banff, Canada.

Daly, A.J. and J. de D. Ortúzar. (1990). "Forecasting and Data Aggregation: Theory and Practice". Traffic Engineering and Control. Dec. 1990. pp. 632-643.

Meurs, H. (1991). "A Panel Data Analysis of Travel Demand". Doctoral dissertation. Univ. of Groningen.

Morikawa, T. (1989). "Incorporating Stated Preference Data in Travel Demand Analysis". Doctoral dissertation, MIT. Cambridge, Mass.

Morikawa, T., Ben-Akiva, M. and McFadden, D. (1990). "Incorporating Psychometric Data in Econometric Demand Models". Banff Invitational Symposium on Consumer Decision Making and Choice Behavior. Banff, Canada.

Widlert, S., M. Bradley, E. Kroes, R. Sheldon, T. Gärling and S.Uhlin (1989). "Preferences for Bus and Underground Services in Stockholm". WCTR World Conference on Transportation Research. Yokohama.