STATED PREFERENCE METHODS – BACK TO THE FUTURE

Eric Kroes
Significance and VU University Amsterdam
Rob Sheldon
Accent

1. INTRODUCTION

In the late 1970s and early 1980s the first Stated Preference (SP) studies started to emerge in transport demand analysis. An early example of such a study was Meyer et al. (1978) describing a mode choice analysis using what they called "functional analysis", a form of SP research.

One of the very first European applications of a SP approach to a transport problem was presented at the PTRC Summer Annual Meeting in "The Use of Conjoint Analysis in Transport Research" by Sheldon and Steer (1982). Since then many SP studies have been carried out, and nowadays SP is one of the accepted standard tools of the transport planner.

This paper looks back at Sheldon and Steer's early SP application more completely reported in Steer, Davies and Gleave (1981), and reviews it in the light of the many lessons that have been learned in the 35 years of SP research that followed. We assess the key elements of the original application, including issues such as the preference model, survey method, experimental design, preference elicitation, coefficient estimation, link to revealed preference behaviour, market simulation and the derivation of demand elasticities. And we compare these with the state-of-practice in 2016.

Finally we draw some conclusions about the quality of the early work by Sheldon and Steer and the practical possibilities for carrying out SP research nowadays with all the up-to-date tools and techniques available to the practitioner. And we point at an area where the development has not been so evident.

2. SHORT DESCRIPTION OF THE 1982 SP STUDY BY SHELDON AND STEER.

The paper by Sheldon and Steer (1982) provides a general introduction into "conjoint analysis" together with a few case studies to illustrate its application to transport research. The methodology description used by Sheldon and Steer (1982) borrows heavily from a classic marketing research paper about conjoint analysis by Green and Srinivasan (1978). The transport case studies they describe were carried out by the authors themselves and are original.

In this paper we focus on the methodology and implementation results for their analysis of demand for Intercity rail travel, a project that was commissioned by the British Railways Board in 1979. The aim of that project was to estimate demand elasticities with respect to service frequency and interchange on Intercity train services in the UK. We describe this study below in terms of its key elements and its main results.

2.1 Number of attributes and levels

Following an extensive in-depth interview programme and a literature review, the so-called "attribute-set and attribute-levels" were determined.

Based upon previous research the authors decided to include four attributes in the SP exercise: travel time, fare, frequency of service and the number of interchanges required. Given the aim to arrive at no more than twelve attribute level combinations, they decided to include three different values for each attribute: the existing situation (as a reference), an improvement and a worsening.

2.2 Preference model

Sheldon and Steer used a classical compensatory preference model belonging to the Multi-Attribute Utility Theory (MAUT) family, using a so-called part-worth function specification:

$$U_k = \sum_i f_i (y_{ik})$$

where:

U_k = the utility for individual k
f_i = the function denoting the part-worth of different levels of y_{ik} for the i-th attribute.

In practice, only three levels of y_{ik} were used, and the part-worths for intermediate y_{ik}'s could be obtained using linear interpolation.

This specification was one of the usual approaches in conjoint analysis in those days, and was related to the coefficient estimation software that Sheldon and Steer used: MONANOVA (Kruskal 1965).

That software estimated eight independent coefficients for each individual respondent, using the nine observations obtained from this individual. It goes without saying that the statistical precision of these estimates was extremely poor. Sheldon and Steer then obtained aggregate results for segments of travellers by taking the mean of their coefficient values, which they called preference weightings.

2.3 Survey method

A key feature of this study was the customisation perspective. Sheldon and Steer developed an elaborate interviewing approach that they called the *Journey Planning Game*. With this they tried to replicate as closely as possible the passenger's choice process when deciding which train they preferred among the available possibilities to travel.

This was consistent with the observation by Fishbein and Ajzen (1975) who had observed that the more specific the context to which the questions related, the more robust the answers given by the respondents.

The Journey Planning Game approach determined how a respondent had made their travel decision and, as said, looked to replicate this process.

A key input was the written timetable – the main source of information in those days – and the authors painstakingly re-wrote these for all the options for all the origins and destinations involved. A lot of Tipp-Ex and ink was got through (younger readers might need to look up what Tipp-Ex was).

The surveys were carried out on train with rail passengers during the course of their journey, using pencil and paper. The surveys were administered by well-trained professional interviewers.

The choice options presented to the respondents for evaluation were customised by the interviewers using the reported travel cost, travel time, train departure times (frequency) and number of interchanges as a reference. The experimental design was then used to create a set of alternative journeys that were all derived from the existing reference journey for each respondent. So an example option card could be:

Journey Planning Game – Example Travel Option Card

- Travel time: *1 hour and 22 minutes* (eg existing time minus 20%))
- Travel cost: *£22* (eg existing cost plus 10%)
- Frequency: *2 services per hour* (eg twice the existing service frequency)
- Interchanges: *1 interchange at Birmingham* (as now)

These characteristics were then presented in timetable format rather than in less meaningful terms like 2 services per hour.

Sheldon and Steer argued that by playing the Journey Planning Game the responses given by the respondents were most realistic and hence valuable, thus eliminating as much as possible any biases in the answers they obtained.

2.4 Experimental design

Sheldon and Steer used four different attributes in their game, each of which could take three values: improvement, existing level, worsening. The existing level was the reference journey as reported by the respondent.

A full factorial design with four attributes at three levels each would give $3^4=81$ different combinations in total.

That was clearly too high a number of options for practical application, so Sheldon and Steer decided to use an orthogonal fractional factorial design consisting of just nine different options. This fractional design was taken from an experimental design cookbook, and allowed for the estimation of all main effects[1], but not for estimation of any of the possible interaction effects[2]. The experimental design was specified as follows:

Fractional factorial 3^4 design

	Attribute 1	Attribute 2	Attribute 3	Attribute 4
• Option 1:	1	1	1	1
• Option 2:	1	2	2	3
• Option 3:	1	3	3	2
• Option 4:	2	1	2	2
• Option 5:	2	2	3	1
• Option 6:	2	3	1	3
• Option 7:	3	1	3	3
• Option 8:	3	2	1	2
• Option 9:	3	3	2	1

This resulted in nine different option cards that were presented for evaluation to each respondent.

2.5 Measurement scale for preference

Sheldon and Steer required the respondents to rank order their nine option cards, indicating which option they preferred most, which they preferred least, and what the order of preference was for all seven intermediate options.

Because the options were presented in the form of little cards, with hand-written values on them, the respondents could actually move their option cards on a table until they were happy with the order. The resulting rank order was then recorded by the interviewers on the survey form, together with the values for each attribute of each card, and coded and data-entried subsequently.

2.6 Coefficient estimation

As already indicated above Sheldon and Steer used MONANOVA to estimate the part-worths, which they also called preference weightings. MONANOVA used the recorded rank order to produce for each respondent a set of twelve weightings, one value for each of the attribute-level combinations.

By relating each coefficient to the coefficient of the reference level, two relative weightings could be determined for each attribute: one for improvement and one for worsening. This made it possible to obtain different

[1] A main effect is the utility of the attribute irrespective of the values of the other attributes.
[2] An interaction effect occurs when the value of one attribute affects the valuation of another attribute; for instance in a very crowded train, where many travelers have no seat, the disutility of travel time will be higher than in a quiet train; so the value of travel time interacts with the level of crowding.

estimation results by the direction of change. Note that Sheldon and Steer did not seem to be aware of the work of Kahneman and Tverski (1979), but most of the estimates they obtained were actually in line with prospect theory.

2.7 Derivation of elasticities

The main objective of the Sheldon and Steer Intercity rail study was to derive service frequency elasticities. So the preference weightings (of undetermined metric) needed to be converted into demand elasticities.

Sheldon and Steer did this in two steps. First they standardised the preference weightings into weighting per unit change, in order to eliminate the influence of the units used for specifying the attributes. Then they determined the ratio between the averaged standardised frequency weightings to the averaged standardised travel time weightings, and multiplied the result with a "known" journey time elasticity (in this case they used a "known" travel time elasticity for improvement of -0.35). This way they derived elasticity values separately for worsenings and improvements of frequency, fare and interchange.

2.8 Link to RP behaviour

In 1982 it was not yet commonly known that in order to prepare realistic travel demand forecasts using SP data, the SP preference utilities (or weightings) should be scaled to an observed RP scale. Morikawa (1989) was the first to systematically research and describe this requirement, which was well after Sheldon and Steer's study. But implicitly, based upon intuition, Sheldon and Steer had already done this when they estimated the frequency elasticities by bringing in the "known" journey time elasticity as a real-life benchmark.

2.9 Market simulation

Sheldon and Steer were not interested in developing explicit demand forecasts. For them it was sufficient to obtain frequency demand elasticity values for a large number of different market segments. They obtained these by averaging the preference weightings for the respondents in the segments of interest before scaling. They assumed that the results of the survey were representative for the target population of the study, and they used a minimum segment size of 30 respondents. No explicit market simulation was carried out, which is understandable because no specification of a realistic travel demand choice set was available or needed (the choice set of nine option cards was an artefact created by the researchers).

3. SUMMARY DESCRIPTION OF THE CURRENT STATE-OF-PRACTICE IN APPLIED SP STUDIES

Since 1982 conjoint analysis studies and stated preference methods in general have become standard tools of the transport planner, and much progress has been made in all their methodological elements. In the following sections we summarise some of the main developments that have taken place and briefly describe the current state-of- practice for applied SP studies.

This is not exhaustive but gives a flavour of the changes that have taken place. We refer to, for instance, Louviere, Hensher and Swait (2000) for a more extensive description.

3.1 Number of attributes and levels

The number of attributes and levels used in SP transport studies varies a lot between different types of studies. In simple value-of-time experiments often only two attributes (time and money) are used, presented to the respondents at a range of different levels going from three to more than ten levels. Other SP studies aiming to find valuations for a larger number of attributes have included many more attributes in a single experiment, often 6 or so, but even up to 10 attributes. And combinations of multiple experiments with one or more common attributes have been used to allow much larger numbers of attributes to be valued.

Also, studies aiming to develop mode choice models have also attempted to capture alternative specific constants as additional variables in the logit model.

So, there is a marked increase in scope of what is now possible cf this original study.

3.2 Preference model

Nowadays in transport SP studies, the use of the binary or multinomial logit (MNL) model is totally dominant, following the widely accepted transition to discrete choice based preference measurement consistent with Random Utility Maximisation (RUM).

This has logically led to the use of discrete choice models with typically linear additive utility functions, representing an assumed compensatory nature of utility. Variations of this model form can also be used to represent non-compensatory preferences, including for instance a lexicographic model where the utility estimate of the most important attribute dominates that of the second most important attribute, and so on. More recently also other decision rules have been proposed such as for instance random regret-minimisation (RRM, see eg. Chorus et al. 2008), but these have not (yet) been widely used in practical SP studies.

Other developments worth noting are the use of covariates to allow for person- and context-specific influences on the preferences, and the use of interaction terms between different attributes. Covariates are often introduced through segmentations in the preference model, or through interactions between explanatory attributes and context descriptors. Sometimes dummy-variables are used.

Essentially the coefficient estimation procedure tends to involve decomposing the average preferences of the respondents *top-down* into values that differ significantly between subpopulations or contexts.

This is vastly different from this early study where preferences were established *bottom-up*, initially at the level of individual respondents and were subsequently grouped into segments and contexts with similar preferences.

3.3 Survey method

The widespread access to and use of the internet since the 1990s and the development of (portable) computer technology have led to substantive changes in the nature of market research.

Whereas in the 1980s almost all surveys used face-to-face, mail out or telephone channels, nowadays a reasonably high proportion of surveys are facilitated through the internet. So these days we have computer assisted personal interviewing (CAPI), computer assisted telephone interviewing (CATI), and computer assisted web based interviewing (CAWI).

These developments mean that the recruitment of respondents is often effected through internet contact, that the interviewing process itself has become computer assisted, and that the data coding and storing process is almost entirely automated. As a consequence the average cost of surveys has become less. And the quality of the data has often improved although without care it can potentially be less good when using commercial panels with respondents using straight-line approaches, for instance, to responding.

In practice, then, survey administration can often still benefit from having interviewers involved, particularly for the more complex applications.

So the array of interview options has vastly increased from those early days.

3.4 Experimental design

The use of choice based experiments has led to different requirements for the experimental design of SP studies. Given that (generic) logit models are by definition specified in terms of utility *differences* between the choice options, the traditional cookbook designs were no longer useable. Louviere et al. (2000) devote an entire chapter to the design of choice experiments, as opposed to the more traditional orthogonal experimental designs used by Sheldon and Steer.

In this context an important distinction needs to be made between experiments for *labelled* (alternative-specific) and *unlabelled* (generic) attributes. For choice experiments with generic attributes only, researchers such as Rose and Bliemer (2009) have developed a family of so-called *efficient* designs that aim to maximise the statistical precision of the resulting model estimations. When some *a-priori* knowledge is available about the approximate coefficient signs and values of the explanatory attributes, optimal designs can be constructed that reduce the sample size required to achieve a certain precision of the estimates. Although this sounds extremely attractive in theory, in practical SP studies in transport these techniques have not yet been widely applied.

This is often partially due to the lack of a-priori knowledge about the coefficients but can also be due to the complexity of the method (although computer tools are available that assist in its execution) and to the fact that many typical transport problems involve and require alternative specific coefficients.

3.5 Measurement scale for preference

Choice-based experiments have replaced the use of rank-ordering and scaling[3] as direct utility assessment methods by discrete choices in a choice set consisting of a small number of realistic choice alternatives.

It was claimed that such a task would be easier to carry out for a respondent as it resembled closely the choice task in real choice behaviour. Another important benefit of this was that the replication of a realistic choice set lead to error variances that could be closer to the error variances in real choice contexts.

3.6 Coefficient estimation

Following the transition to choice based preference measurement logit has become the dominant coefficient estimation method. Methods used nowadays in practical SP studies range from simple binary or multinomial logit to nested logit models with multiple scaling factors. Other more complicated structures do exist, but are rarely used in practice.

The coefficient estimation models mentioned produce coefficients for groups of respondents rather than the individual-specific coefficients as produced by MONANOVA. But by using segment-specific covariates and/or distributed parameters, differences in preferences between individual respondents can be captured as well.

Another issue in coefficient estimation is the fact that in SP multiple responses are obtained from the same respondent. This has consequences for the error-term that are ignored in standard logit model estimation, and generally leads to underestimated standard deviations of the estimated coefficients. There are standard methods to deal with this, such as jack-knife and bootstrap methods and more recently also error-components logit and latent class logit. However, the vast majority of applied studies still ignores the issue, relying on the assumption that the estimated coefficients are unbiased.

3.7 Derivation of elasticities

The use of the logit model has opened two different ways of deriving demand elasticities using SP data:

- The first method is simply based upon the use of the standard demand elasticity (*point* elasticity) formula for the logit model with linear additive utility function:

[3] With scaling we mean questions which ask respondents to express their strength of preference using for instance a 5-point scale (1-2-3-4-5) or a 7-point scale (1-2-3-4-5-6-7).

$\varepsilon_i = \alpha_i x_i (1-P_m)$,

where:
ε_i = (demand) elasticity for attribute i
α_i = utility coefficient for attribute i
x_i = (mean) value for attribute i
P_m = probability of choosing mode m.

- The second method is based upon the use of market simulation using the demand utility function, which is applied to a representative sample of individuals. The model is applied to forecast demand twice, once with the existing values for all attributes and once with one of the attribute values changed by for instance 10%. Then the ratio is determined between the predicted change in demand and the assumed change in attribute value (in this example 10%). The *arc* elasticity value for attribute I is then simply this ratio. The methods of market simulation are briefly described in section 3.9 below.

In the first case, the use of P_m which is the actual, observed market share of mode m, automatically brings in the "scale" of real choice behaviour. In the second case, the "scale" (or error variance) of the logit model needs to be explicitly adjusted to reflect the scale of true Revealed Preference behaviour. Morikawa (1989) was the first to research and describe this requirement. Nowadays it is common practice to carry our combined SP and RP estimation, using data sources of both kinds, so that the SP data can be used to derive accurate trade-off coefficients and the RP data are used to provide a realistic application scale of the final demand model.

3.8 Link to RP behaviour

In sections 2.7 and 3.7 we have already seen that it is necessary to scale the preference utilities in order to obtain utilities that can be used in a realistic demand forecasting model. Sheldon and Steer's method of using a "known" demand elasticity as a benchmark is no longer applied, and has most often been replaced by a more sophisticated combined SP/RP estimation as described in the previous paragraph.

3.9 Market simulation

Demand forecasts for entire markets can be derived using SP logit models. This can be done in very much the same way as it is done in disaggregate transport models. Several methods do exist for this, with the main ones being sample enumeration, the uses of synthetic samples and micro simulation of entire populations. We briefly discuss each of these below.

Sample enumeration involves the use of the same sample of trips that was used for the estimation of the demand model, as the basis for an application of the model under changed conditions. The model is applied in a probabilistic way, i.e. the predicted probabilities for each choice alternative are computed and added across all members of the sample. This way the results are exactly reproducible: repeating the simulation will give exactly the same result.

Synthetic samples can be used when the original estimation sample of trips is too small, or too restricted in terms of circumstances it represents. Here an artificial pseudo sample of individuals and contexts is created that can be used to apply the model to. The model is applied in probabilistic way, just like the sample enumeration application.

Micro simulation of the entire population typically works slightly different. Here the demand model is applied to each member of the population, but now in a deterministic way. The model is used to predict the probabilities for each choice alternative, and then Monte Carlo simulation is used to determine one quasi "chosen alternative". Only that single chosen alternative is recorded for each member of the population. Obviously repeating the same process with a different starting values of the random generator used for the Monte Carlo simulator will produce different "chosen alternatives" for many population members. But adding these across all members should give an outcome that is close to the previous one, depending on the size of the simulated population (with very large populations the differences will be very small).

4. COMPARISON BETWEEN SHELDON AND STEER 1982 AND TYPICAL 2016 SP STUDY

When we compare the study of Sheldon and Steer with a typical SP study in 2016 we see that a lot of progress has been made, in almost all key methodological elements. This is illustrated in Table 1, where a summary comparison is presented.

Table 1: Comparison of methodological elements between Sheldon and Steer (1982) and a typical 2016 SP application

Element	Sheldon and Steer 1982	Typical 2016 SP application
1 Number of attributes and levels	4 attributes at 3 levels each	Highly variable, 2-10 attributes at 2-10 or more levels
2 Preference model	Part-worth function coefficients	RUM with linear additive utility function
3 Survey method	Face-to-face interviews inside trains, customised travel options on cards	Computer Assisted Web based Interview (CAWI), customised choice options
4 Experimental design	Fractional factorial orthogonal design	Random draws from full factorial design, or efficient design (eg. D-efficient)
5 Preference elicitation	Rank-ordering of 9	Repeated discrete

	choice options	choices among 2, 3, 4 or more choice options
6 Coefficient estimation	Monotonic Analysis of Variance at individual level	MNL model, with or without nesting
7 Derivation of elasticities	Ratio of preference weights related to "known" reference elasticity	Direct from coefficients or using market simulation, after RP scaling
8 Link to RP behaviour	Implicitly through reference elasticity	Explicitly by using (1-P) or RP scaling or joint SP/RP estimation
9 Market simulation	Not applicable	Sample enumeration or application to synthetic sample after RP scaling

At the same time it can be noted that Sheldon and Steer did it "right" in several very important ways:

- They used a highly customised survey approach, with a lot of attention being given to presenting realistic and comprehensible alternatives to each individual traveller;
- They used four attributes at three levels each, in combination with an orthogonal design with only one dominant alternative, which worked very well for their rank-ordering task;
- They estimated separate coefficients for worsening and improvement relative to a known existing reference situation, which is consistent with prospect theory as developed by Kahneman and Tverski (1979);
- They did not use a realistic travel choice set in combination with a discrete choice, but the rank-ordering approach that they used was consistent with their choice set and was analysed correctly using MONANOVA as a dedicated analysis method;
- They used the ratio of preference weights to infer demand elasticities, by relating these to a "known" reference elasticity, which was consistent with their linear additive part worth function and which provided their implicit link and scaling to RP behaviour.

Some of this was deliberate, based upon a very careful preparation and an extensive qualitative research programme preceding their SP surveys. But some was based upon intuitive guesses that turned out to be correct in retrospect. Looking back, we can only say that this "correctness" is likely to have contributed to the substantial popularity that SP methods have received in subsequent years.

One key area that has largely been ignored is that of customisation through *gamification*. The Journey Planning Game worked well but it was arduous to compile and this effort has been evidently lacking in many subsequent studies. Now, this is being put back onto the agenda and a good example of a modern day elaboration is provided in a paper being presented at this conference on Friday entitled 'The impact of rail fares complexity on demand' (Metcalfe *et al.* 2016). In this example a modern day ticket sales channel is replicated and used as the basis for a complex stated preference experiment.

5. CONCLUSION

In their first major European SP study in transport, Sheldon and Steer (1982) did a very good job in many ways: their early work was based upon good thinking, a very careful preparation and correct intuition rather than established theory. In those days there were no examples that could be copied, no standard formats that could be adapted. Their transport survey application needed to be created almost from scratch.

Since then a lot of progress has been made, in all key dimensions of the method. In this paper we presented the key elements of their early study, and contrasted these with the current state-of-practice in applied SP studies.

In 2016 the transport researcher has much more knowledge about these methodological elements, based upon research into what should be done and what not. And also there are nowadays several software tools that can be used to conduct an SP study in a correct and efficient way addressing important methodological elements such as:

- The use of computer assisted interviewing techniques;
- The creation of experimental plans both for orthogonal and efficient designs;
- The estimation of advanced discrete choice models, including SP/RP scaling;
- The derivation of demand elasticities and policy simulation.

One very important reason for the success of Sheldon and Steer has been their extensive use of *customisation*, in what they called the *Journey Planning Game*. By speaking the language of the respondents, and by presenting travel choices that closely resembled the real context of the travellers, they were able to generate meaningful preference information that could be turned into sensible numerical information. This is an area that has been neglected in recent years and deserves to be explored with greater creativity.

REFERENCES

Chorus, C.G., Arentze, T.A. and Timmermans, H.J.P. (2008): A random regret-minimization model of travel choice, *Transportation research part B*, Vol. 42(1), pp.1-18.

Fishbein, M.A. and Ajzen, I. (1975): Belief, attitude, intention and behavior: an introduction to theory and research, Reading, MA, Addison Wesley.

Green, P.E. and Srinivasan, V. (1978): Conjoint Analysis in Consumer Research: Issues and Outlook, *Journal of Consumer Research,* Vol. 5, No. 2 (Sep., 1978), pp. 103-123.

Kahneman, D. and Tverski, A. (1979): Prospect Theory: An Analysis of Decision Under Risk, *Econometrica*, National Emergency Training Center.

Kruskal, J.B. (1965): Analysis of factorial experiments by estimating monotone transformations of the data, *Journal of the Royal Statistical Society*, Series B, 27, pp.251-263.

Louviere, J.J., Hensher, D.A. and Swait, J.D. (2000): Stated Choice Methods: Analysis and Applications, Cambridge University Press.

Metcalfe, P., Heywood, C. and Sheldon, R. (2016): The impact of rail fares complexity on demand, paper to be presented at ETC 2016, Barcelona.

Meyer, R.J., Levin, I.P. and Louviere, J.J. (1978): Functional analysis of Mode Choice, Transportation Research Record, Nr. 673, pp 1-7.

Morikawa, T. (1989): Incorporating stated preference data in travel demand analysis, PhD dissertation, Department of Civil Engineering, MIT.

Rose, John M. and Bliemer, Michiel C. J. (2009): Constructing Efficient Stated Choice Experimental Designs, *Transport Reviews*, 29 (5): pp 587–617.

Sheldon, R. and Steer, J. (1982): The Use of Conjoint Analysis in Transport Research, PTRC Summer Annual Meeting, PTRC, London.

Steer, Davies & Gleave Ltd (1981): Final report on research into Elasticity of Demand in Respect of Service Frequency and Through Trains.