Designing and Implementing a Stated Preference (Conjoint) Exercise

FINAL REPORT

Copies of this document can be downloaded from the website below.

BNR Consulting, 22nd November 1993

Author: Dr. Lawrence Harrell MCIT, MMRS
7 Austin Avenue, Poole BH14 8HD ENGLAND
Tel./fax.: +44 (0)1202 299491
Web: http://www.demandanalysis.co.uk, Email: see website

1. Introduction:

This paper provides a simple overview of the design and implementation of a Stated Preference (Conjoint) exercise to estimate the value of attributes. It is assumed that the reader has a general idea of what Conjoint Analysis (also termed Stated Preference) is and how it operates.

2. Design:

2.1. Experimental Design:

Initially the number of attributes and levels has to be decided upon. Generally all designs will include money (in terms of fare/price) to enable the values of other attributes to be expressed in some meaningful way. After fares other attributes are included that are of interest to the researcher.

There are some limitations on the attributes that can be included in the exercise. As the number of attributes and levels are increased the number of replications needed in the experimental design increases rapidly. This can be seen by looking at one of the experimental design catalogues or by generating designs using the computer software. Common sense tells us that the more information you require the greater the amount of data you have to collect.

Conjoint exercises generally involve each individual performing all replications of the design. There is a limit to the number and complexity of designs that can be handled by a respondent - the limiting factors are time and information overload. The latter occurs where an individual is presented with more information than can be processed in their short-term memory. Individuals response to this situation can vary from refusing to complete the exercise to ignoring attributes (not always in a systematic fashion).

There are some in the industry who believe that, as in the real world there are a very large number of attributes associated with any product, it is perfectly reasonable to present individuals with a large number of them in such a hypothetical exercise. This approach is typically called "full profile conjoint". There are a number of potential flaws in this argument. Firstly it is known that when information overload occurs respondents ignore attributes, usually these are the least significant ones. Using a full profile approach to estimate values for secondary (often intangible) attributes is therefore very dangerous. Furthermore in the real world the plethora of attributes are not presented in such an artificial way i.e. by textual descriptions with one or two illustrations. A full profile exercise therefore required far more cognitive ability (in a short space on time) than a real world situation.

For a Conjoint exercise a minimum of three attributes are required in order to hide the exact purpose of the exercise from the respondent. How far the number of attributes are increased above three is a matter of debate. It is generally recognised (for example, Journal of Transport Economics and Policy, January 1988) that eight is a maximum.

Much depends on the implementation of the exercise. If it were to be given to commuters for self-completion before they boarded their train - even four attributes may be too many. Generally any design that requires an individual to complete more than about twelve complex replications without some form of reward (i.e. money, free travel etc.) is perhaps pushing respondents too far. Related to this is the possibility of non-response bias caused by over-complex survey instruments. It is possible that a certain type of person may be more likely to refuse to compete such surveys, these could be those with a high value of time or those who are least educated - both of these are important segments in any exercise.

Generally the more time and the greater the supervision available to the respondent, the greater the complexity of the design that can be given. The decision to use rating, rating or choice (or pseudo choice) approaches will also affect the difficulty of the exercise for respondents.

2.2. Intangible Attributes:

Designing an experiment that attempts to incorporate intangible attributes requires more care. Intangibles (like noise, ride, comfort etc.) do not generally have a readily comprehensible measurement scale. To use such attributes in a Conjoint exercise it is therefore necessary to produce some unambiguous form of measurement that is understood by the respondent, while still meaning something to the engineer. If such a scale can be successfully produced the respondents will understand the exercise and the results can be used to make policy judgements when the results have been analysed.

For a detailed discussion of valuing intangibles using Conjoint analysis a useful reference is Harrell (1990).

2.3. Interactions:

Unfortunately some attributes are not completely independent of each other. A good example would be the value of the view from a hotel room and the value of the level of the room. The higher the level the more important a view - these attributes are not independent and if this relationship is ignored parameter estimates are likely to be invalid and not very significant.

Floor	View	Interaction
1	1	1
2	2	4
3	1	3
4	2	3
1	1	8
2	2	4
3	1	3
4	2	8

Interactions are usually considered by specifying the number to be included when choosing the experimental design and adding into the data an additional attribute that is the product of the levels of the related attributes. For example hotel floors may be 1,2,3,4 and view coded as 1 (poor view) 2 (good view). The possible interaction term would be coded as below

Interactions would generally only be an issue when considering intangible attributes. Ride quality and noise would be an example. it should be noted that many commercial Conjoint packages have trouble dealing with interaction effects.

2.4. Disaggregate Designs, Aggregate Designs and Bridging:

Two types of designs can be produced: aggregate (replications are divided between a group) or disaggregate (where each individual does all replications). The disaggregate approaches gain as each respondent only has to complete a small number of replications, this means that a more complex experimental design can be used with many more attributes considered.

However there are some significant drawbacks with disaggregate designs. It is not possible to calibrate at the individual level. This means that it is not possible to check each respondent's logic. It can also be very difficult (or even impossible) to identify market segments after the data has been collected. For example trying to estimate the value of time for male business air passengers may not be effective unless the combination of the replications considered by these individuals are orthogonal. Generally it is recognised that the "noise" in such aggregate models mean that parameters produced by such models are less robust than those generated by disaggregate models.

One compromise that is often used is a bridging design where the experimental design is split between say three or four people. This approach is more difficult to calibrate and does not allow a check of respondents logic - though it does allow a greater number of attributes to be considered.

A final and more recently favoured way of increasing the number of attributes without increasing the effort required by each respondents is to generate a series of different designs each with a common attributes (usually money). Obviously such an approach increases the sample size and care must be taken to ensure that the various samples are matched as closely as possible. This approach allows checks on respondents to be made and reduces the noise in estimating parameters.

2.5. Recording of Respondents Preferences:

2.5.1. General:

After presenting respondents with a series of options, their reactions to the various scenarios have to be recorded in some way to allow the effect of the various attributes on product preferences to be estimated. Conjoint analysis is based on Microeconomic Utility theory which states that individuals are rational and will pursue the course of action that maximises their utility. It follows from this that the response scales used in conjoint analysis represent some proxy for utility.

There are three groups of ways of measuring respondents preferences. The simplest two for the analyst are ranking and rating. Developing an exercise based on choices between options is considerably more complex for the analyst (although easier for the respondent).

2.5.2. Ranking:

Respondents can rank the options presented to them. For example if there were ten cards, preferences would be coded from one to ten. To make the exercise easier ties can be allowed for. With ranking exercises respondents generally have more difficulty considering cards in the middle (in terms of preference). Allowing ties between middle ranked cards can therefore greatly speed up the exercise. Thus with ten cards a ranking may be: 1,2,3,4,4,4,7,8,9,10. A ranking exercise produces a great deal of information as respondents are comparing each card with every other card. Ranking exercises are also very simple to design and analyse, though it has recently become less fashionable.

One drawback with ranking is that any exercise with more than say ten cards can become very difficult for the respondent.

Ranking based approaches are also criticised as being unrealistic: in the real world people choose to purchase something - they don't rank things in order of preference. One advantage of this unrealism is that it can be used to estimate parameters in sensitive situations (for example when there is a likelihood of policy response bias) as this exercise makes the choices less apparent to the respondent.

Ranking approaches can be improved by using a little imagination. Ranking can be made easier by asking respondents to initially select their best and worst options. These are then taken away and the respondent is asked to select their best and worst from the remaining cards. This process continues until there are no cards remaining. Using this variation of ranking makes the exercise closer to a choice exercise.

2.5.3. Rating:

With a rating approach each option is given a score of (say) between one and ten. This approach is the easiest to design and analyse and allows the use of regression for calibration without breaking statistical assumptions. But once again there are problems with a conventional rating technique.

Rating a series of options in sequence can result in inconsistent scaling by respondents. A respondent when confronted with the first card does not know what the other cards will be like and therefore is not sure how to rate each option. Until the respondent has seen all cards he is therefore unable to effectively rate them. This suggests that the most effective approach is to present the respondent will all options at once, making it similar to a ranking approach - though more difficult. When all options are presented together the respondent must not only put the options in rank order, but also consider the size of the difference between options.

2.6. Complex Recording of Respondent's Preferences - Choice:

2.6.1. General Approach:

In a regular choice exercise respondents are presented with two or more options and asked to choose between them. In a real choice exercise no scaling is used, as the choice is a simple "A" or "B" scenario. Such a binary choice makes the data unsuited for analysis by regression (which is not able to successfully deal with categorical dependent variables) meaning that expensive specialist tools like ALOGIT or BLOGIT have to be used to estimate a utility function based on the maximum likelihood technique.

Practitioners argue that a simple choice like this is the most realistic and some variation of the choice approach is perhaps the most commonly used in transport planning.

An often used variation on the choice approach also incorporates elements of a rating. Respondents are presented with two options and are commonly given a five point response scale. This scale usually has options of: definitely choose A, probably choose A, indifferent, probably choose B, definitely choose B. This scale is usually interpreted by the analyst as expressing the probability of the respondent using mode B. It is frequently coded as 0.1, 0.3, 0.5, 0.7, 0.9.

This dependant variable is sometimes transformed for the analysis to produce a logit curve to represent the choice between modes.

$$\text{Probability of choosing Option A} = \frac{\exp^{-}\text{(Utility of Option A)}}{\exp^{-}\text{(Utility of Option A)} + \exp^{-}\text{(Utility of Option B)}}$$

A logit curve is "S" shaped illustrating the fact that individuals are more sensitive to changes in products the closer (in terms of generalised cost) alternatives are to each other. In other words people are more responsive to product changes in a more competitive environment. This transformation is done by using the "log odds ratio". Applying this approach the dependent variable becomes -2.20, -0.85, 0.00, 0.85, 2.20.

$$\text{Natural Log of:} \quad \frac{\text{Probability of Selecting Option A}}{\text{Probability of Selecting Option B}}$$

2.6.2. Implications for Design:

In this paper only pairwise choices will be considered to reduce complexity. In a choice based exercise each replication from the experimental design represents both options. There a number of ways of approaching the design of pairwise Conjoint exercises.

Initially it seems that representing two alternatives on each option would require a design with many more attributes, thus greatly increasing the number of replications required in the experiment. For example consider an option comparing two trains (A and B) where three attributes (say: fare, time and frequency) are being used to distinguish between them. This suggests that six attributes are required in the exercise (fare A, fare B, in-vehicle time A, in-vehicle time B, frequency A and frequency B).

Fortunately there is a simple way to side-step these problems - the attributes can be considered in terms of differences between A and B. We might thus consider three levels of fare ($1 more, $2 more and $3 more). It is also possible to mix these approaches and express some attributes as differences and others in terms of levels.

When using a "differences" based design care must be taken to ensure that the attributes of the two alternatives are comparable. For example, when comparing different modes of transport it may be wise to consider in-vehicle time by car as a different attribute to in-vehicle time by train.

2.7. Other Issues and Conclusions:

2.7.1. Presentation of Cards:

Care should be taken with order effects. These occur when respondents who complete a series of similar tasks are affected in some way by performing these tasks. For example increased practice may make a respondent better at the task. Continued questioning about a subject may also make respondents more sensitive to the subject. The repetitive nature of Conjoint exercises could generate such difficulties.

The conventional way to handle these effects is to randomise the order in which stimuli are presented to respondents. The order of presentation is often randomised by presenting each option on a separate card, these are shuffled between respondents.

2.7.2. Customisation of Exercise:

It is accepted that the further the exercise departs from the experience of the respondent the less valid are the results. It is therefore important to only ask respondents about what is familiar to them. If a new product or service is to be investigated, it is therefore important to explain its characteristics carefully in a non-leading way. Otherwise a valid response is unlikely to be elicited.

There are two issues that could limit the desire to make the exercise as realistic as possible. The problems of information overload associated with full-profile conjoint have been discussed earlier. The second problem concerns policy response bias. The latter may be an issue when considering the introduction of new services. People are usually keen to see an improved service and have been known to suggest that they would use a service that they have no intention of using - in an attempt to encourage its provision. Great care therefore has to be taken when using Conjoint analysis to predict demand for new services.

2.7.3. Choice of Approach:

It should be clear from this discussion that the design of Conjoint experiments is subject to a great deal of uncertainty - there is no single "correct" way to do it. Generally a choice based approach will be the easiest for the respondent, though a modified ranking approach can be useful in a simple study or in sensitive situations. Smaller designs will generally be more effective than larger ones, as they result in greater comprehension and produce a higher response rate.

2.7.4. Simulation of Data and Piloting:

The problems discussed so far emphasise the importance of testing the design before the full scale survey. Ideally two forms of testing should be used - simulation of data and pilot testing. Data can be simulated on a spreadsheet using random numbers. The simulated data can then be analysed to see if the design is able to pull out meaningful coefficients. For example if the value of time is expected to be $20 p/hour responses could be created for people having values between $10 and $30 p/hour.

If attribute levels are not chosen carefully, the effect of an attributes may be swamped by the others. For example comparing a $100 change in the single fare with a 5 minute saving in journey time is unlikely to yield meaningful results. The simulation of data would reveal this problem before implementation.

Even if the design has been carefully tested in this way, it is still crucial that the exercise is fully piloted. This pilot process must take place in the field with "real" respondents and the data must be analysed - again to see if meaningful coefficients are produced. Failure to follow these steps can lead to some very expensive mistakes!

References:

Harrell L. The Effects of Intangible Product Attributes on Rail Passenger Demand (with Particular Reference to Ride Quality). Ph.D. thesis (March 1990), Centre for Logistics and Transport, Cranfield Institute of Technology, Bedford MK43 0AL, England.

Various Authors in The Journal of Transport Economics and Policy - Stated Preference Issue (January 1990).