What information do the error terms hide in Choice Modelling environmental valuation studies?

John Rolfe
Central Queensland University
P.O. Box 197
Emerald, 4720
Ph. 0749 822 904

Jeff Bennett
University of New South Wales

Jordan Louviere
University of Sydney

Abstract.

Choice modelling (CM) is a developing non-market valuation technique that provides a rich data set to the analyst. In addition to parameter estimates for the influence of attributes, labels and respondent characteristics, models also provide information about, (and are sensitive to), the error terms implicit in model parameters.

The error terms provide information about the appropriate model form to be analysed in at least three important areas. First, specifications of nested logit models, which are useful for improving model validity, are based on allowing correlation to occur within (but not between) the error terms for choices grouped together in nests.

Second, the error terms for separate, but similar CM experiments can be used to generate ratios of the scale parameters that are confounded with individual model parameters. This enables the results from different models to be compared, and identifies choice models that have larger amounts of inherent variability.

Third, by identifying the error terms associated with successive choices in CM experiments, it is possible to search for learning and fatigue effects that might be displayed by respondents. A series of these disaggregation exercises have been performed on CM experiments dealing with rainforest valuation to determine where learning and fatigue effects might be present.

Key Words: Choice modelling, random utility models, learning effects.

1.0 Introduction.

Choice Modelling (CM) is a non-market valuation technique that provides researchers and policy makers with an alternative to the contingent valuation method (CVM). While the CM technique has some potential strengths in the estimation of non-use values, there are only a limited number of applications in this field to date, and a detailed understanding of how best to apply the technique is still developing.

One of the potential issues in applying CM is the extent to which learning and fatigue effects are present in the responses to the choice sets. A CM experiment usually involves respondents being asked to complete a sequence of choice sets. There is a wide variety of ways of designing these choices, so that the complexity of the choices being modelled, the number of tradeoffs present within a choice set, and the number of choice sets that may be given to respondents are all likely to vary.

The presence of learning and/or fatigue effects would imply that the accuracy of CM results is being compromised because some of responses are inconsistent, perhaps reflecting a higher random error component in the choices. To generate more accurate models of choice, researchers generally try to minimise learning and fatigue effects. A better understanding of their existence would aid in this regard.

Researchers are generally well aware of the potential problems associated with learning and fatigue effects. Applications of valuation techniques usually reflect standard accepted practices of survey design and application. However, applications of CM to environment issues where non-use values are being estimated generally involve unfamiliar choices. One question of interest then is whether the involvement of abstract environmental issues is more likely to stimulate learning and fatigue effects than would the topics of more conventional marketing studies.

The presence of learning effects in an experiment is of interest for two further important reasons. The first of these is validation. If people take some time in a hypothetical setting to develop consistent choice criteria, especially when the case study may be unfamiliar and complex to them, then it is likely that learning effects may also be present in real-life choices[1]. This gives another dimension to issues of validation between the hypothetical (stated preference) experiments, and the real-life (revealed preference) data, usually gained from market transaction information.

The second reason why potential learning effects are of particular interest in environmental valuation issues is that they have some implications for the accuracy of the CVM technique. In contrast to CM, CVM essentially asks for a response to a single tradeoff between the status quo and some development or conservation position. If the first response is less accurate because of learning effects, this would suggest that the applications of the CVM may be less accurate than applications of CM in the same situation.

Some understanding of the role of error terms in CM results is the pathway to analysing potential learning and fatigue effects. Choice randomness can be identified both between and within respondents in CM studies, thus enabling choice behaviour to be disaggregated in different ways.

These issues are explored in this paper with reference to two CM experiments that involved the assessment of preservation values for tropical rainforests. A discussion of the reasons

[1] In real life situations people are likely to evaluate the transaction costs of gaining better information against the benefits flowing from improved choices.

why learning and fatigue effects might exist in relation to environmental amenities and issues is the focus of the next section of the paper. In section three, the hypotheses of interest are developed, together with an outline of the statistical tests that are used on the available data sets. In section four the experiments of interest, together with the results of the statistical tests are reported, and section five concludes the paper.

2.0 Learning and fatigue effects.

Choice modelling has its roots in the marketing, psychology and economics literature, and has been derived from the family of conjoint analysis techniques commonly employed in marketing. Since the introduction of the CM variants by Louviere and Woodworth (1983), which are also referred to by terms such as experimental choice (eg Adamowicz, Boxall, Williams, and Louviere 1998), CM has developed into a technique that is appropriate for economic valuation tasks. It is not very dissimilar in theoretical underpinnings from referendum CVM models.

Many of the implicit guides for applying CM derive from standard practices in marketing techniques and the intuition and experience of CM practitioners. The application of CM surveys often involves a compromise between the amount of data required for statistical accuracy, (particularly when specialised models are being estimated), and the choice burdens that can realistically be imposed on potential respondents (Louviere 1988, Batsell and Louviere 1991). While experimental designs can be partitioned through some blocking mechanism into smaller split-samples to reduce the number of choice sets for each individual respondent, practical limits about the complexity of each choice set also exist (Batsell and Louviere 1991).

However, there is very little theory or empirical evidence to inform ‘best practice’ in CM applications (Brazell and Louviere 1998), and wide variation in practice about the format and complexity of experiments that are presented to respondents in survey form (Carson et al 1994). This is in part because of the developing nature of the CM technique, and the focus of researchers to date on other methodological issues. It may also be because ‘best practice’ varies widely according to the case study at hand, and not enough detailed knowledge is yet available to inform researchers about appropriate limits in various settings.

A wide diversity in approaches from other fields about where appropriate limits lie for the application of choice questionnaires has previously been noted by Louviere, Oppewal, Timmermans and Thomas (1993). In particular, some approaches to surveying from the psychology and sociology fields involve tasks that appear to be more complex and onerous than is commonly believed possible by economic practitioners (Brazell and Louviere 1994). This suggests that more empirical work is needed to determine appropriate limits to the setting of complex tasks in CM experiments.

Learning and fatigue effects may be thought of as functions of complexity and burden aspects of an experiment. The overall size of a CM experiment can be conveniently disaggregated into profile, choice set and survey length aspects. The experiment operates by asking respondents to identify a preferred option among a set of profiles, which are short descriptors of the potential scenarios available. This choice task is repeated across other sets of profiles so that the researcher builds up a rich data set of choice information and is then able to identify statistical models of choice.

The profiles used in experiments are comprised of a number of attributes relevant to the situation of interest, each of which can vary across a set number of levels. The profile differences are generated in some systematic manner by varying the level that each attribute can take in a particular profile. Large numbers of attributes and/or levels create complex profiles that survey respondents may have trouble in visualising and understanding. This the complexity burden that may be contributed by individual profiles. This means that while it is possible to have large numbers of attributes or levels or profiles within a choice set, it is not practical to have large numbers of each simultaneously (Batsell and Louviere 1991, Carson et al 1994).

A typical choice set for respondents will involve two or more profiles, together with a 'No Choice' or other opt-out option. Respondents are requested to indicate their preferred choice from the options given. An additional complexity burden may thus be associated with the size of the choice sets, and may be directly related to the number of profiles within each set.

Respondents are typically asked to complete a series of these choice sets. The overall burden then can be thought of as a function of the complexity of individual choices, together with the total number of choice sets that respondents are expected to complete. Of course, other factors, such as the experience and capability of respondents, the choice patterns employed and the burden of the framing stages and other experiment duties may also be important.

Complexity and burden can be more interrelated than this analysis might imply. Respondents to CM surveys are often cautious about the repetitive nature of the choice sets. Once they realise that the choice sets are very 'repetitive', they typically look carefully for any hidden patterns or 'tricks' in the tasks, and try to ensure that they are consistent in the way that they make their choices (Blamey, Rolfe, Bennett and Morrison 1997). This suggests that a different form of cognitive burden is added to subsequent choice sets once the first one has been completed.

On the other hand, the complexity of repeated choices may be diminishing as respondents become familiar with the inherent tradeoffs and the situation facing them. In this situation, the overall burden of a choice experiment may not be very strongly related to the number of choice sets because the cognitive burden of each choice continues to fall.

Learning effects.

Louviere (1988) and Carson et al (1994) recommend that one or more practice choice sets be included at the beginning of a CM exercise to minimise learning effects. Alternatively, Louviere (1988) suggests that respondents be given example sets and provided with the bounds on the range of attribute combinations. Louviere (1988) notes that Myer (1977) found that the decision strategies of survey respondents stabilised after three choice sets in a three attribute experiment. This implies that more than three practice sets are needed for more complex designs. Brazell and Louviere (1998) report that Johnson and Orme (1996) found that reliability of parameter estimates increased through 20 choice sets.

However, Brazell and Louviere (1998) found that choice consistency appeared to improve for the first 36 choice sets, but that the consistency began to decline past some optimum number of choice sets (approximately 50). However, the differences between their segmented samples were not statistically significant, indicating that while that there was some visual evidence of learning effects, they were unlikely to be impacting on parameter vectors.[2].

Learning effects may conveniently thought of as two separate effects. In the first, respondents learn how to make more accurate choices as they learn about the task at hand. This will be reflected as a change in the underlying parameters of the choice models being estimated, meaning that different welfare estimates can be gained as respondents change criteria.

[2] Visual effects refers to the comparison between choice consistency and survey length.

In the second, the underlying choice criteria remain consistent, but the variability attached to choices changes as respondents learn to do the task at hand. These types of learning effects will be represented as reductions in the variability of choices as respondents move through the sequence of choice sets. As the task moves from being novel to familiar, the random component of utility is likely to fall.

Preferences for non-market environmental goods may include larger random components than preferences for normal market goods. This is because:

- individuals are generally not experienced in making choices involving environmental goods,
- individuals generally have low levels of knowledge of environmental goods as compared to environmental goods, and
- there is rarely information available about the types of choices that other people have made.

While many learning effects seem to be confined to reductions in the random error component and have little influence on the underlying choice parameters (Brazell and Louviere 1998), there has been little analysis of learning effects in applications of the CM technique to environmental issues. To explore this issue in relations to our experiments estimating the preservation values of tropical rainforests, our hypothesis about learning effects is that:

H1: Choice consistency is initially positively correlated with the number of choice sets.

To test for learning effects, it is helpful to disaggregate this hypothesis to reflect the two separate effects outlined above. The first of these hypothesises that learning effects are reflected in changes to model parameters, while the second hypothesis that learning effects are related to changes in the distribution of error terms.

H1a: There will be some relationship between changes in model parameters and the progressive number of choice sets.

H1b: The variance of the error term distribution will fall in relation to the progressive number of choice sets.

Fatigue effects.

If the burden of performing a choice experiment rises throughout the course of the experiment, then the reliability of responses may fall as respondents become fatigued. CM practitioners have traditionally tried to kept experiments ‘short and simple’ to minimise complexity and fatigue problems (Louviere 1988), although the evidence from other disciplines is that fatigue effects may be minimal (Brazell and Louviere 1998). In this case, opportunities may exist to design larger experiments and improve analytical.

The evidence about fatigue effects is mixed because fatigue is not only a function of survey length, but the cognitive burden within choice sets and the additional choice burden involved in a survey instrument. This relationship is well-recognised within the literature (eg Louviere 1988, Carson et al 1994), but few empirical tests are available. Most efforts to test for fatigue have focused on survey length. For example, Brazell and Louviere (1998) tested whether the length of a CM experiment was inversely related to response rates and found only weak support for the hypothesis. Their results imply that it may be possible to use much longer survey instruments in CM experiments.

Swait and Adamowicz (1996) suggested that survey length would be directly related to the task burden facing respondents, and therefore inversely proportional to response rates.

Bradley and Daly (1994) analysed choice consistency as a function of survey length in an attempt to isolate respondent fatigue. They found that consistency fell as respondents moved through their choice tasks, suggesting that respondents simplified decision rules, paid less attention and/or responded non-systematically in order to complete the task. However, there may also be some contrary end effects, where respondents increase their effort and dilengence towards the completion of a task (Brazell and Louviere 1998).

If fatigue effects do occur, they might become manifest in two important ways. The first of these is through decreased reliability of parameter estimates, while the second is through increased variability in their responses. If respondents become fatigued or disinterested as they move through a CM experiment, then it is probable that they would become less discriminating in their responses. This can be represented through the second hypothesis:

H2: Choice consistency is negatively correlated with the position of choice sets in an experiment.

As with learning effects, this can be conveniently restated as two sub-hypotheses, being

H2a: There will be some relationship between changes in model parameters and the progressive number of choice sets, and,

H2b: The variance of the error term distribution will in relation to the progressive number of choice sets.

3.0 Testing the Hypotheses.

In order to test the hypotheses, a framework has to be adopted where the differences in model forms and parameters can be compared. The random utility approach underlying the CM technique is appropriate for this purpose. This approach describes the utility of a choice as being comprised of a systematic (explainable) component and an error (unexplainable) component. The following equation formalises the basic relationship where V_{ij} represents the measurable component of utility and e_{ij} captures the effect of unobserved and omitted influences on choice[3].

$$U_{ij} = V_{ij} + e_{ij} \qquad \text{... (1)}$$

The systematic component of choice can be disaggregated further, as in the following example where utility is held to be a function of the characteristics of the relevant good (represented by Z_{ij}) and the characteristics of the individual (represented by S_i), together with the error term.

$$U_{ij} = V(Z_{ij}, S_i) + e_{ij} \qquad \text{... (2)}$$

Choices made between alternatives will be a function of the probability that the utility associated with a particular option (*j*) is higher than for alternatives, as in (3).

$$P_{ij} = \text{Prob}(V_{ij} + e_{ij} > V_{ih} + e_{ih}) \quad \text{for all h in Choice set C, } j \neq h \quad \text{... (3)}$$

[3] More formally, an stochastic error term is associated with the utility of choice to represent the effect of random response shocks, while another error term is associated with the influence of unobserved characteristics. The latter error term is unique to each individual respondent, and may be reduced by introducing heterogeneity into choice models. Alberini, Kanninen and Carson (1997) demonstrate this in relation to CVM. In this analysis the two error components are confounded.

The parameters for the relationship can be introduced by assuming that the relationship between utility and characteristics follows a linear in the parameters and variables function, and by assuming that the error terms are distributed according to a double log (Gumbel) distribution, the choice probabilities have a convenient closed-form solution known as the multinominal logit model (MNL) (McFadden 1974). The MNL model is generally preferred because it is computationally easier to use (Stern 1997), and takes the general form:

$$P_{ij} = \exp(\lambda V_{ij})/\sum \exp(\lambda V_{ih}) \qquad \text{(for all h in choice set C)} \qquad \ldots (4)$$

where λ represents a scale parameter which is commonly normalised to 1 for any particular data set. The MNL model generates results for a conditional indirect utility function of the form:

$$V_{ij} = \lambda(\beta + \beta_1 Z_1 + \beta_2 Z_2 + \ldots.\ \beta_n Z_n\ +\ \beta_a S_1 + \beta_b S_2 + \ldots.\ \beta_z S_n) \qquad \ldots (5)$$

where β is the constant terms, and β_1 to β_n and β_a to β_z are the vector of coefficients attached to the vector of attributes (V) that influence utility. The constant term β can be partitioned into alternate specific constants (ASCs) that are unique for each of the alternatives that are considered in the choice sets. These ASCs capture the influence on choice of unobserved attributes relative to specific alternatives.

The "error" variability in a CM experiment arises from a number of sources, including variability within individuals (over choice sets) and variance between individuals (true heterogeneity). This variability can be attached to the individual components of choice (ie associated with beta values of attributes), and/or attached to the alternatives included within choice sets. Variability is driven by large number of factors, such as the prior knowledge of respondents and the presentation of the survey instrument, and thus may be expected to differ between case studies and data sets.

The scale parameter is inversely proportion to the variance of the error term, as in the following where μ^2 is equal to the variance of the error term[4]:

$$\lambda = \pi^2/6\mu^2 \qquad \ldots (6)$$

The scale parameter cannot be identified in a specific model because the error terms are confounded with the vector of utility parameters where β_k (the vector of utility parameters) can be more accurately represented as $\lambda\beta_k$ (Ben-Akiva and Lerman 1985, Swait and Louviere 1993). This means that results of different models are not directly comparable, as differences in variance affect the magnitude of model parameters. However, it is possible to estimate the ratio of scale parameters between different data sets, and hence, test the equality of parameter estimates (Swait and Louviere 1993, Louviere 1994).

One way of determining trends in model parameters is to examine the part-worths that may be estimated from MNL models for individual choice sets. The data are disaggregated according to the choice sets, so that individual MNL models can be estimated for each successive choice that respondents have made. Welfare estimates can be estimated from MNL models through the use of the following formula:

[4] The error terms (e_i) are distributed according to the extreme value distribution with a mean of zero and a variance of μ^2. This variance of the error terms is impossible to isolate from a single data set because changes in β and μ^2 such that the ratio β/μ^2 remains constant have no effect on probability (Stern 1997).

$$CV = -1/\alpha[\ln\sum \exp v_{io} - \ln\sum \exp v_{i1}] \quad \text{... (7)}$$

where CV is the compensating variation welfare measure, α is the marginal utility of income and V_{i0} and V_{i1} represent indirect utility functions before and after the change under consideration., the marginal value of a change within a single attribute can be represented as a ratio of coefficients, where equation 7 reduces further to:

$$W = -1x\beta_{attribute}/\beta_{money} \quad \text{... (8)}$$

This part-worth formula effectively provides the marginal rate of substitution between income change and the attribute in question. Because the scale parameters are cancelled out in the estimation of a part-worth formula, part-worths between different models can thus be directly estimated. A comparison of the part-worths for the ASC's and the attributes across different choice sets will help to identify where learning and fatigue effects are occurring, and give some idea of the direction of the change in values. A graph of these marginal values provides some visual indication of where differences in underlying parameters may exist. If the part-worths are constant across choice sets, then this would indicate that no learning or fatigue effects are present.

To determine whether differences in model parameters are significant, and the direction of change in error terms, the Swait-Louviere test is employed[5]. This is the pathway to test formally H1 and H2. A likelihood ratio test is used to determine whether the underlying parameters of models estimated from the different choice sets are equivalent once differences in scale parameters have been accounted for. This explicitly tests H1a and H2a. The ratio of scale parameters will also help to determine which models have more variance associated with them, and thus help to determine the order of choice sets with higher deterministic outcomes. This will allow us to test H1b and H2b.

The likelihood ratio test operates by stacking the two data sets to be compared, and then varying one of the datasets by a scalar factor such that the log-likelihood value of the MNL model estimated from the joint data set is maximised. The scalar factor that maximises the log-likelihood value is the effective ratio of the scalar factors associated with the two data sets. When no scale difference exists between the data sets, the scalar factor will equal one. If the ratio of λ_1 to λ_2 is less than one, this implies that data set 1 has more variability than data set 2, and vice versa.

The likelihood ratio test proposed by Swait and Louviere (1993) is used to test whether the difference in parameter vectors for the datasets is significant. The test statistic is given by:

$$LR = -2[LogL_{\lambda 1/2} - (LogL_1 + LogL_2)] \quad \text{...(9)}$$

where $LogL_{\lambda 1/2}$ is the log-likelihood of the MNL model estimated from the combined data set, and $LogL_1$ and $LogL_2$ are the log-likelihoods calculated from the respective individual models. The test statistic is asymptotically distributed as a chi-square with $[k_{\lambda 1/2} - (k_1 + k_2) + 1]$ degrees of freedom, where $k_{\lambda 1/2}$, k_1 , and k_2 represent the number of parameters estimated in the separate models (Blamey, Bennett, Morrison, Louviere and Rolfe 1997).

If significant learning and/or fatigue effects are present, then parameter differences between the models estimated for successive choice sets should be significant, and the ratio of the scale factors will give some indication about which data sets have higher error terms.

---

[5] Tests for parameter equivalence can also be performed from the part-worth data through the Krinsky and Robb (1986) procedure. The Swait-Louviere proceedure is preferable though because of the additional information about the scale parameters that is generated.

4.0 Performing the experiments.

The two experiments described in this paper were designed to estimate the willingness-to-pay that Australians might hold for the preservation of rainforests in several locations, including the country of particular interest, Vanuatu[6]. In both experiments, the profiles for the proposed conservation areas were described according to common seven attributes, being:

- location,
- area
- rarity
- potential to visit,
- effect on local populations,
- existence of special features,
- cost to respondent of preservation.

The profiles are effectively described by location, three environmental attributes (area, rarity, special features), and three socio-economic factors (potential to visit, effect on local populations, and cost to respondent). This means that the potential tradeoffs made by respondents may have been quite complex. The cost attribute was set up as a donation mechanism to make preservation options in overseas countries more realistic.

One survey is generic in design with only two profiles included in each choice set. This meant that the location variable was randomised across the choice sets in accordance with the experimental design, and only two locations were available in each choice set from the pool of eight. In contrast, the other experiment had labelled options, where each of six locations were offered consistently across the choice sets. There was some slight difference in the number of levels between the two choice sets in order to meet some experimental design parameters, but the "upper" and "lower" levels (eg $5 and $50) are common across the two experiments. Samples of the choice sets used are presented in the appendices.

The problems of complexity and cognitive burden were recognised early in the design phase, and several measures were taken in the survey instrument to minimise potential biases. First, a parsimonious number of attributes and levels was chosen to minimise the size of profiles in the choice sets. Second, very little definitional information was given to respondents about rainforests and possible conservation sites. Instead, a series of preliminary questions in the survey instrument was designed to encourage respondents to focus on rainforest conservation and potential conservation tradeoffs.

The series of preliminary questions (common to both surveys) was also designed to minimise potential learning effects. Respondents were asked to rank different environmental and social tradeoffs in terms of importance, give an indication of how many times they had visited rainforests in Australia and overseas, and provide some feedback on their purchases of some environmentally friendly goods. In one question they were asked to indicate their potential choice between a shopping basket of normal goods and a shopping basket of environmentally friendly goods with a certain weekly (and annual) cost difference. In another question, they were asked to indicate their potential choice between a shopping basket of environmentally friendly goods and support for a rainforest conservation project.

One purpose of these questions was to remind respondents that there was often a monetary cost in supporting environmental protection, that the cost was borne by individuals in society, and that people had a great deal of choice in supporting environmental programs. As well,

[6] These experiments have been described in more detail in Rolfe, Bennett and Louviere (1997, 1998).

the questions were designed to give respondents some familiarity with choosing between "bundles" of goods, especially when these 'bundles' were both potentially appealing to respondents.

There were no practice or demonstration sets included in the CM sets within the questionnaire, and no final questions to evaluate the difficulty or comprehension that respondents experienced with the choice sets. In part, the exclusions of these learning aids and feedback mechanisms was prompted by a desire to keep the surveys as short as possible and minimise fatigue effects.

The surveys were performed in Brisbane, with 105 respondents completing the generic experiment survey in July 1995 and 200 respondents completing the labelled experiment survey in May 1996. Other results of these survey have been reported in Rolfe, Bennett and Louviere (1997, 1998).

5.0 Analysis and discussion.

The first step in analysing the survey data was to group the data according to choice sets. Thus all the first choice sets that respondents complete go into one data set, and so on. Basic MNL models are calculated from each choice set, and part-worth comparisons used to give some idea of how the results compare between choice sets[7]. The same data sets were then used for performance of the Swait-Louviere tests.

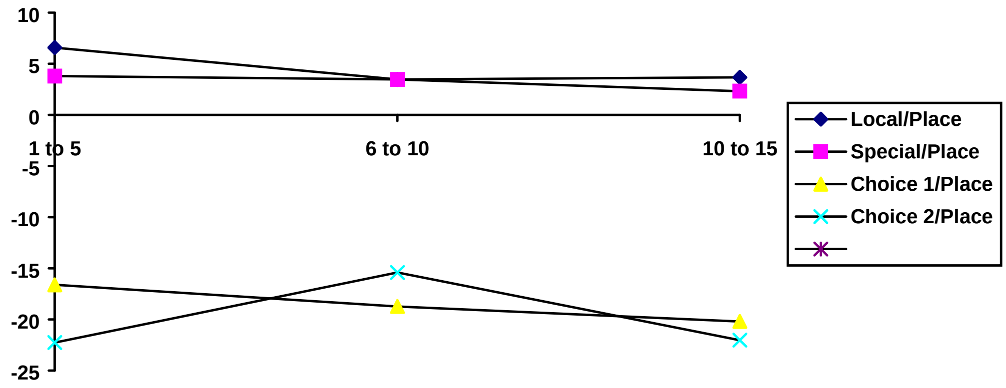
It proved difficult to estimate models with high explanatory power from the generic model data set because of the survey design employed and the limited number of respondents. To address this, the data for several choice sets was grouped in an effort to improve statistical efficiency. This was not particularly desirable, because grouping may disguise learning and fatigue effects between choice sets. It was only when five choice sets (from a total of 16) were grouped into data sets that reasonably robust models of choice could be developed.

The results of some part-worths calculated from these models are graphed in Figure 1[8], as well as being reported in more detail in Appendix 1. They show that there is some slight downwards trend in partworths over the choice sets, as well as slight variation in that trend. This suggests that some effect, perhaps a learning effect, caused the relative values to decline over the course of the experiment.

Figure 1. Part-worths in successive choice sets - generic model experiment.

[7] Models were also estimated with socio-economic variables, but little difference in outcomes was observed. For clarity, the analysis in this paper is based on basic MNL models.

[8] Place was used as a base for calculating the part-worths, as one of the coefficients for price was not significant.



The data from the labelled experiment proved to have high explanatory power, and basic MNL models could be calculated from the individual choice sets. Some indication of trends can be gained from Figure 2, which plots the part-worths of successive choice set models. The table of part-worths for the different choice sets is reported in more detail in Appendix 2.

Figure 2. Part-worths in successive choice sets - labelled experiment.

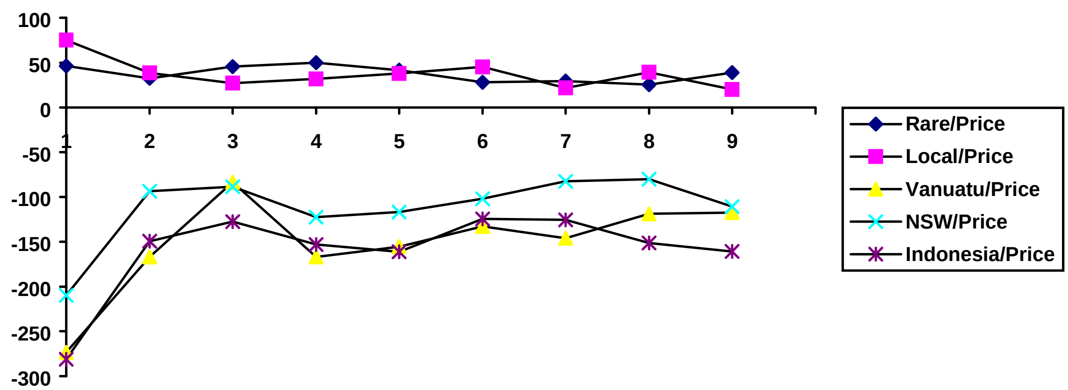


Figure 2 indicates how the vector of parameters is changing between successive choice sets, and thus is indicating where the learning and fatigue effects defined in H1a and H2a may be located[9]. The figure shows that some learning effects may be apparent, especially in the difference between the first and second choice sets. There is also some variation apparent in the value estimation process in the first four choice sets as compared to the later five sets. However, there does not appear to be any long term trend in part-worth value changes that could be used to identify more substantial learning or fatigue effects. Thus the visual evidence from the part-worth data from this experiment is that there may be some learning effects in the first few data sets, that the learning effects are not necessarily consistent, and that there do not appear to be longer term learning and fatigue effects.The first step in a more rigorous test of whether learning and fatigue effects exist is to see whether significant differences in model parameters exist once scale parameter effects have been isolated. It is possible that as respondents learn (and tire), their underlying values driving their responses changes. This first test of the equality of parameters for separate choice set models then is a

[9] These graphs do not identify whether the changes in part-worths are significant.

test of the possible effects of learning and fatigue symptoms, and is performed with a Swait-Louviere test.

The performance of a test can be illustrated by comparing the models that can be generated from the first and second choice sets in the second experiment, where the graphical representations indicated that potential learning effects were strongest. A basic MNL model is calculated for each data set in turn, and then the two sets are combined and the scalar factor is used to isolate the joint choice set with the highest log-likelihood value. The value of the scalar factor that was identified between these tow choice sets was 0.94. This meant that model fit was maximised when the variables in the 2nd choice set were rescaled by 0.94.

The results of the three models are shown in Table 1. Application of the log-likelihood values calculated generates a statistic of 8.7878. This compares to the test Chi-square statistic of 19.675 at a 95% confidence interval and 11 degrees of freedom. As a result, the hypothesis that differences in parameter estimates are significant once scale parameter differences have been removed can be rejected. While it is possible that learning effects are present in terms of choice variability, it does not appear that any learning effects are impacting on model parameters.

Table 1. Swait-Louviere test between two choice sets in labelled experiment.

Variable	Choice set 1	Choice set 2	Choice set 1&2
Area	0.00007	0.00004	0.00006
Rare	0.450	0.486*	0.593
Visits	0.174*	0.001	0.129*
Local	0.747	0.573	0.651
Special	0.010*	0.241	0.140
Price	-0.010	-0.014	-0.010
Vanuatu	-2.726	-2.500	-2.762
Far North Qld	-0.907	-0.839	-1.002
NSW	-2.091	-1.402	-1.881
South America	-2.501	-2.066	-2.455
South east Qld	-1.031	-1.139	-1.275
Indonesia	-2.801	-2.233	-2.599
Log-Likelihood	-315.3489	-341.2412	-660.9840

This procedure has been performed across the data for the two experiments. The results of the tests in the generic model experiment are reported in Table 2. These show that once scale parameter effects have been accounted for, some significant differences in model parameters still exist. As discussed below, this suggests that some learning effects may be present in the experiment.

Table 2. Swait-Louviere tests across choice set groups in generic experiment.

	ChoiceSet 1-5	ChoiceSet 6-10	ChoiceSet 11-15	Joint 1-5/6-10	Joint 1-5/11-15
LogL	-483.5593	-476.2880	-485.5999	-976.6573	-983.6289
Lambda				0.99	0.98
Swait-Louviere statistic				47.563	43.482
Chi-Square statistic				21.026	21.026

Significant?	Yes	Yes

The tests have also been performed across the choice sets in the labelled experiment. For consistency, the median choice set (number five) has been chosen as the base, and comparisons made with the other choice sets in turn. The results are summarised in Table 3, and show that the hypothesis of significant differences can be rejected in each case compared to a Chi-square statistic of 19.675.

Table 3. Swait-Louviere tests on choice sets - labelled experiment.

Choice set	Lambda	LogL (i/5)	LogL (i)	LogL (5)	S-L test	Signif	Variance
1	1.05	-674.723	-315.349	-351.737	15.2754	No	1.492
2	0.95	-695.285	-341.241	-351.737	4.6136	No	1.8226
3	1.02	-664.88	-308.326	-351.737	9.6338	No	1.5811
4	1.06	-697.658	-338.737	-351.737	14.3696	No	1.4639
5	1.00						1.6540
6	0.83	-686.678	-329.574	-351.737	10.7342	No	2.3878
7	0.89	-691.243	-334.468	-351.737	10.0752	No	2.0766
8	0.95	-683.261	-325.665	-351.737	11.718	No	1.8226
9	0.93	-696.732	-336.286	-351.737	16.5986	No	1.9018

These results produce conflicting evidence about whether learning effects might exist in environmental CM experiments. While it is clear that learning effects do not impact on the parameter vectors in the labelled experiment, there does appear to be some impact in the generic experiment. This is unexpected, because the logic underlying the fourth hypothesis is that choice modelling sets in the generic experiment would be easier than in the labelled experiment, and thus less likely to be associated with learning effects.

The inclusion of socio-economic variables in the choice models makes little difference to results. When seven socio-economic variables were included in the MNL model specification, the resulting model differences were still significant (Table 4). This suggests that the apparent learning effects are being driven by more than heterogeneity.

Table 4. Swait-Louviere tests across choice set groups in generic experiment with socio-economic variables included within model.

	ChoiceSet 1-5	ChoiceSet 6-10	ChoiceSet 11-15	Joint 1-5/6-10	Joint 1-5/11-15
LogL	-470.0201	-472.7836	-462.6940	-963.241	-958.9839
Lambda				0.98	0.98
Swait-Louviere statistic				40.8746	52.5396
Chi-Square statistic				36.415	36.415
Significant?				Yes	Yes

A further explanation of the differences in model results is that orthogonality of the choice sets may not be consistent. The experimental designs used generate choice tradeoffs that are orthogonal across the full experiment, but may not be orthogonal across individual choice sets. The reduction in orthogonality would be limited in the labelled model because of the six profiles included within each choice set, but may be more problematic in the case of the generic model. As a result, it appears likely that both H1a and H1b can be rejected.

The second stage in analysing the choice set models is to compare the scale parameter ratios that were generated between the different choice sets. Where the parameter vectors are equivalent, the variance attached to the different choice sets can be compared with some confidence. This variance is expressed in terms of the lambda coefficients reported in tables 2 and 3. The scale parameter of a data set is inversely proportional to the variance of the error term, as in equation 6.

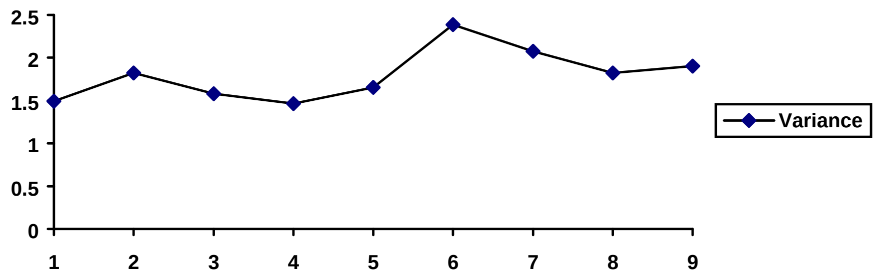
If the ratio of scale parameters λ_1 to λ_2 is less than one, this implies that the expression $\pi^2/6\mu^2$ is higher for data set two. By extrapolation then, the variance of the error terms is lower for date set two than it is for date set one. The scale parameter ratios in Table 2 and Table 3 thus give some indication of how the distribution of error terms compares between data sets. If the first (or base) data set has less variation than the other date set, the ratio of scale parameters will be more than one, and vice versa.

The ratio of scale parameters from the choice blocks in the generic experiment is slightly less than one where the first choice block was used as the base. This indicates that choice variability has tended to fall marginally across the choice sets. This suggests that the parameter differences indicated in the Swait-Louviere tests are consistent with learning effects rather than fatigue effects. As respondents have progressed through the choice sets, the variability of their choices has fallen and they have adjusted their responses in line with their learnt information.

The variability information available from the labelled experiment is more complex. The fifth choice set has been used as a base, so the results indicate that the second, sixth, seventh, eighth and ninth have all had greater choice variability. For convenience, the scale parameter ratios have been converted into variances (see equation 10), and these are reported in the final column of Table 3, and illustrated in Figure 3 below.

The movement in the variance of error terms is not very compatible with the hypothesis of learning effects being present. The increase in variance between the first and second choice set is consistent with some focus group feedback where respondents complete the first choice set easily, but then are confused by the similarity of the next choice set. Many respondents report being puzzled by the repetitive nature of the choice sets, and being anxious about the possibility that they may be giving contradictory answers. One possible explanation of the variance data shown in Figure 3 is that respondents are finding the second choice set difficult to do because they are not used to the repetitive nature of the questioning, and the latter half of the choice sets difficult because they are finding it very difficult to process the choices and use decision rules that are fully consistent.

Figure 3. Variance of the error distribution across choice sets - labelled experiment.



These results suggest that there is little support for H1b to be accepted. While the results of the generic survey suggest that it is possible for learning effects to be present in an experiment, the results of the labelled experiment indicate that they are not automatically associated with choice experiments, and do not necessarily follow easily defined trends.

There is also little support for H2b to be accepted. While there is some limited evidence that choice variance can rise in situations where respondents are aware of complex tradeoffs, there is no evidence that this followed any long term trend that would be expected in a fatigue effect. Considering that the lengths of the surveys reported here are modest, this result is not surprising.

Conclusions.

Learning and fatigue effects in applications of the Choice Modelling technique are a potential concern to researchers because they may reduce the accuracy of subsequent model estimates. In this paper, tests to identify learning and fatigue effects from two CM applications to a non-market valuation issue have been reported. To facilitate these tests, it has been helpful to think about learning and fatigue effects as two separate influences on MNL models.

The first way that learning and fatigue effects may influence models is to change the values of parameters, so that different welfare estimates are ultimately generated. Learning effects should change model parameters closer to 'true' parameters, while fatigue effects may move them away. The second way that learning and fatigue effects may influence models is through the random error component of choices. Here the underlying parameters of choice remain unchanged, but the random components (error terms) may fluctuate. It would be expected that learning effects would be associated with a reduction in the variance of error terms, while fatigue effects might be associated with an increase.

A comparison of the part-worths generated from the MNL models estimated from the choice data indicated that there may be some learning effects but few fatigue effects associated with changes in the underlying model parameters. This was tested more formally with the Swait-Louviere tests. For the generic data set, the parameter differences did appear significant, but this may have been a function of orthogonality problems in the data set. For the labelled model, no significant differences in model parameters could be distinguished. This indicated that the hypotheses about learning and fatigue effects impacting on model parameters could be rejected.

There was also little evidence available from an analysis of the distribution of error terms that major learning and fatigue effects were present. There was some evidence that choice variance rose between the first and second choice set, consistent with respondents becoming aware that the choice tasks were repetitive. After this, choice variance decreased for three choice sets, consistent with some learning effects. Choice variance rose in the second half of the choice tasks, perhaps indicating that respondents lost interest in the task, but fell towards the last choice sets, consistent with some end effect where respondents put in an extra effort to finish the tasks. This pattern is consistent with results reported by Brazell and Louviere (1998).

The results reported above indicate that learning and fatigue induced choice variability is not directly related to the progression of choice sets, but is likely to be the result of a more complex interaction between choice complexity and respondent behaviour. As a result, there would appear to be no simple relationship available between the dimensions of a CM experiment and the variability of responses. However, further understanding of the relationship will continue to be important in attempts to develop more robust models of choice.

The results of the tests reported here are generally positive for practitioners of non-market valuation. They show that learning and fatigue effects do not appear to be serious impediments to the successful application of CM experiments to non-use values, and that the types of framing exercises employed in the experiments reported here have been successful in introducing respondents to the concepts and tradeoffs of interest.

Further, the implications for the CVM are that responses to a single (or first) choice tradeoff are likely to be accurate. It appears that while there may be some variation in parameter estimates, these are unlikely to be significantly different from the 'true' vector of parameters. There was also little evidence that the first choice set had higher variability than subsequent choices. This indicates that learning and fatigue effects are not automatically associated with the CM technique and the CVM, and that standard design and survey techniques may be very successful in minimising their potential impact.

Acknowledgements.

The research reported in this paper has been supported by the Australian Centre for International Agricultural Research (ACIAR) in Project 9020, *The socio-economic assessment of native forest preservation proposals in Vanuatu: implications for forest management*.

REFERENCES.

Adamowicz, W., Boxall, P., Williams, M. and Louviere, J.J. 1998 "Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation", *American Journal of Agricultural Economics*, 80:64-75.

Alberini, A., Kanninen, B. and Carson, R.T. 1997 "Dichotomous Choice Contingent Valuation Data", *Land Economics*, 73(3):309-324.

Batsell, R.R. and Louviere, J.J. 1991 "Experimental Analysis of Choice", *Marketing Letters*, 2(3):199-214.

Ben-Akiva, M.E. and Lerman, S.R. 1985 *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge.

Blamey, R.K., Bennett, J.W., Louviere, J.J., Morrison, M.D. and Rolfe, J.C. 1997 "The Use of Causal Heuristics in Environmental Choice Modelling Studies", paper presented to the conference of the Australian and New Zealand Society for Ecological Economics, 17-20th of November, Melbourne.

Blamey, R.K., Rolfe, J.C., Bennett, J.W. and Morrison, M.D. 1997 *Environmental Choice Modeling: Issues and quantitative insights*, Choice Modelling Research Report No 4, University College, The University of New South Wales, Canberra.

Bradley, M. and Daly, A. 1994 "Use of logit scaling approach to test for rank-order and fatigue effects in stated preference data", *Transportation*, 21(2):167-184.

Brazell, J.D. and Louviere, J.J. 1998 *Length Effects in Conjoint Choice Experiments and Surveys: An Explanation Based on Cumulative Cognitive Burdens*, unpublished paper, Department of Management, University of Sydney.

Carson, R.T., Louviere, J.J., Anderson, D.A., Arabie, P., Bunch, D.S., Hensher, D.A., Johson, R.M., Kuhfeld, S.F., Steinberg, D., Swait, J., Timmermans, H. and Wiley, J.B. 1994 "Experimental Analysis of Choice", *Marketing Letters*, 5(4):351-368.

Johnson, R.M. and Orme, B.K. 1996 "How many questions should you ask in choice-based conjoint studies", paper presented to the American Association's Advanced Research Techniques (ART) Forum, Beaver Creek, Colorado.

Krinsky, I. and Robb, A. 1986 "Approximating the Statistical Properties of Elasticities", *Review of Economics and Statistics*, 68:715-719.

Louviere, J.J. 1988 *Analysing Decision Making: Metric Conjoint Analysis*, Sage Publishing, Beverly Hills.

Louviere, J J 1994 "Conjoint Analysis", in R Bagozzi (ed) *Handbook of Marketing Research*, Blackwell, Oxford.

Louviere, J.J., Oppewal, H., Timmermans, H. and Thomas, T 1993 "Handling large numbers of attributes in conjoing applications: Who says existing techniques can't be applied? But if you want an alternative, how about hierarchal choice experiments", paper presented to the American Association's Advanced Research Techniques (ART) Forum, Beaver Creek, Colorado.

Louviere, J.J. and Woodworth, G. 1983 "Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data", *Journal of Marketing Research*, 20:350-367.

McFadden, D. 1987 "Regression-Based Specification Tests for the Multinomial Logit Model", *Journal of Econometrics*, 34:63-82.

Myer, R.J. 1977 “An experimental analysis of student apartment selection decisions under uncertainty”, *Special Issue on Human Judgement and Spatial Behaviour: Great Plains-Rocky Mountain Geographical Journal* 6:30-38.

Rolfe, J.C., Bennett, J.W. and Louviere, J.J. 1997 “Choice Modelling - An Exercise in Environmental Valuation”, paper presented to the 41st annual conference of the Australian Agricultural and Resource Economics Society, Gold Coast, January.

Rolfe, J.C., Bennett, J.W. and Louviere, J.J. 1998 “Framing Issues in the Application of the Choice Modelling Technique to International Conservation Proposals”, paper presented to the 42nd annual conference of the Australian Agricultural and Resource Economics Society, Armidale, January.

Stern, S. 1997 “Simulation-Based Estimation”, *Journal of Economic Literature*, 35:2006-2039.

Swait, J. and Adamowicz, W. 1996 “The Effect of Choice Environment and Task Demands on Consumer Behaviour: Discriminating between Contribution and Confusion”, working paper, University of Florida, Gainsford.

Swait, J. and Louviere, J.J. 1993 “The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models”, *Journal of Marketing Research*, 30:305-314.

Appendix 1. Part-worths by choice sets for Generic experiment.

Choice Set	1-5	6-10	11-15
Area	0.001215	0.000126*	0.00082
Rare	1.606944*	5.174843	7.503279
Visits	2.696181	2.466667	0.244262*
Local	6.592014	3.483019	3.683607
Special	3.788194	3.480503	2.311475
Price	-0.159722	0.04025	-0.214754
Choice 1	-16.6285	-18.727	-20.218
Choice 2	-22.2569	-15.4226	-22.0148

Note: * denotes insigificant at the 10% level of one of the coefficients involved.

Appendix 2. Part-worths by choice set for Labelled experiment.

Choice set	1	2	3	4	5	6	7	8	9
Area	0.007	0.0032	0.0029	0.0046	0.0023*	0.0026*	0.0008*	0.0026	0.0042
Rare	46.192	32.473	45.554	49.991	41.414	28.077	29.471	25.351	38.742
Visits	17.52*	0.0852*	-7.472*	17.815	6.6214*	19.548*	17.407	8.2456*	15.786
Local	74.975	38.348	26.901	31.921	37.803	45.097	21.799	39.254	19.874
Special	1.0335*	16.134	15.377	2.8477*	15.607	-4.51*	12.804	8.4211	18.994
Van	-273.5	-167	-83.68	-166.9	-155.5	-132.8	-146	-118.9	-117.4
FNQ	-91.03	-56.05	-80.99	-83.44	-80.92	-41.87*	-61.38	-72.37	-115.7
NSW	-209.8	-93.78	-88.68	-122.5	-116.8	-101.9	-82.54	-80.26	-110.7
SAM	-251.1	-138.2	-109.3	-168.9	-129.5	-140.6	-124.9	-145.6	-190.6
SEQ	-103.4	-76.22	-68.87	-58.94	-105.8	-112.3	-54.5	-63.6	-96.23
INDO	-281.1	-149.4	-127.6	-153	-161.3	-124.5	-125.4	-151.3	-161

Note: * denotes insigificant at the 10% level of one of the coefficients involved.

Appendix 3 A sample choice set from the generic Choice Modelling experiment.

SCENARIO CHOICE 1

Option 1 ***500 hectares in Far North Queensland***	***Option 2*** ***10,000 hectares in Papua New Guinea***
Extremely rare	Fairly rare
Easy to visit with full facilities	Visits allowed, but difficult access and poor facilities
No local people	Protection of rainforest means local people will be better off
Special plants and animals	Special landscapes
$50.00 donation required	$10.00 donation required

Please indicate your preference:
(Tick one)

☐ ***Option 1***

☐ ***Option 2***

☐ ***I would not support either option***

Appendix 4. A sample choice set from the labelled Choice Modelling experiment

Option 1 - Vanuatu
• 10,000 hectares
• Not rare at all
• No visits allowed
• No local people
• Special landscapes as well as plants and animals
$5 donation required

Option 2 - Far North Queensland
• 10,000 hectares
• Not rare at all
• Visits possible but moderate access and few facilities
• Protection of rainforest means local people will be worse off
• Special landscapes as well as plants and animals
$5 donation required

Option 3 - Northern NSW
• 100 hectares
• Not rare at all
• Visits possible but moderate access and few facilities
• Protection of rainforest means local people will be worse off
• Special landscapes as well as plants and animals
$10 donation required

Option 4 - South America
• 1,000 hectares
• Extremely rare
• Visits possible but moderate access and few facilities
• No local people
• No special features
$50 donation required

Option 5 - South East Queensland
• 100 hectares
• Fairly rare
• Easy to visit with full facilities
• No local people
• Special landscapes as well as plants and animals
$5 donation required

Option 6 - Indonesia
• 10,000 hectares
• Not rare at all
• Visits possible but moderate access and few facilities
• Protection of rainforest means local people will be better off
• Special landscapes
$10 donation required

Please indicate preference: (Tick one)

☐ 1	***Option 1***	☐ 2	***Option 2***
☐ 3	***Option 3***	☐ 4	***Option 4***
☐ 5	***Option 5***	☐ 6	***Option 6***

7 ***I would not support any option***