Efficiency benefits of choice model experimental design updating: a case study

Geoffrey Kerr, Lincoln University **kerr@lincoln.ac.nz**

Basil Sharp, University of Auckland **b.sharp@auckland.ac.nz**

Presented to the 53rd annual Australian Agricultural and Resource Economics Society conference; Cairns, Queensland. 10-13 February 2009

Abstract

Efficient experimental designs offer the potential to reduce confidence intervals for parameters of interest in choice models, or to reduce required sample sizes. C-efficiency recognises the salience of willingness to pay estimates rather than utility function parameters. This study reports on a choice model application that incorporated updated statistical designs based on initial responses in order to maximise C-efficiency. The revised design delivered significant improvements.

Keywords: experimental design, choice experiment, efficiency

1. Introduction

Experimental design plays an important role in choice modelling because inappropriate designs may result in unidentifiable models or produce biased parameter estimates (Louviere *et al*., 2000). Inefficient experimental designs fail to capture the fullest extent of information from survey participants, resulting in parameter estimate variances larger than potentially achievable with any given sample size. D-efficiency is the most common approach to measuring efficiency of experimental designs (Ferrini and Scarpa, 2007). D-efficient designs minimise D-error, which is defined as (Scarpa and Rose, 2008):

$$\text{D-error} = [\mathrm{Det}(\Omega(\boldsymbol{\beta}, \mathbf{x}_{sj})]^{1/K}$$

Ω is the asymptotic variance-covariance matrix for the design ($\mathbf{x}_{sj}$) with utility function coefficients $\boldsymbol{\beta}$. K is the number of coefficients estimated. Identification of a D-efficient design entails selection of $\mathbf{x}_{sj}$ which minimises D-error for expected $\boldsymbol{\beta}$. Alternatively, A-efficiency minimises the trace of the asymptotic variance-covariance matrix, which minimises aggregate parameter variances, but may produce very large covariances (Scarpa and Rose, 2008). The main purpose of valuation studies is not estimation of the parameters in the utility function, *per se*, but estimation of willingness to pay (WTP) for environmental attributes. Kanninen (1993) developed designs for contingent valuation studies to minimise the variance in WTP estimates. Similarly, Kessels *et al*. (2006) have proposed G- and V-optimality based on minimisation of maximum and average choice prediction variances. Recently, Scarpa and Rose (2008) have developed choice experiment design strategies to minimise variance in WTP (C-efficiency). They used hypothetical simulations to illustrate the advantages of designing choice experiments in order to maximise C-efficiency, rather than approaches based on D-efficiency and other efficiency criteria.

Some approaches to efficient design assume that all the coefficients are zero. Clearly, such designs fail to utilise all available information in situations in which coefficients are non-zero. Utility function coefficients in non-market valuation studies are expected to be non-zero. For example, environmental protection is a good, pollution is a bad and the marginal utility of money is expected to be positive. In such cases, efficient designs rely upon prior knowledge of the coefficient vector. Such knowledge can come from theory, information obtained from stakeholders during study design and pre-testing, or from sequential data collection. The latter approach uses information obtained in early applications to update the experimental design using either coefficient vector point estimates or using Bayesian updating to account for uncertainty in the coefficient vector.

This study empirically estimates efficiency gains derived by following the C_P design procedure developed by Scarpa and Rose (2008). C_P denotes that the procedure maximises efficiency using point estimates of the coefficient vector rather than the Bayesian approach (C_b). The next section describes the methods used. The results of using a sequential improvement in design are presented in section three. The paper concludes with a discussion of the results and suggestions for further research.

2. Methods

The utility function coefficients and elements of the asymptotic variance covariance matrix can be used to derive confidence intervals for WTP and the sample size required at any desired level of accuracy for any particular WTP value. WTP for attribute i is:

$$WTP_i = -\alpha\beta^{-1}$$

α and β are utility function coefficients for attribute i and cost respectively. Following Scarpa and Rose (2008) the variance of mean WTP may be estimated as:

$$Var(WTP_i) \approx \beta^{-2}(Var(\alpha) - 2\alpha\beta^{-1}Cov(\alpha,\beta) + \alpha^2\beta^{-2}Var(\beta))$$

If one replicate of the experimental design is used to generate an estimate of the asymptotic variance covariance matrix, it is possible to generate a t-score for a single replication of the experiment:

$$t_{i,N=1} = WTP_i\ (Var(WTP_i))^{-0.5}$$

The sample size necessary for mean WTP_i to be significantly different from zero at the 5% significance level is then:

$$N_i = t_{0.05}^{\ 2}\ Var(WTP_i)\ WTP^{-2} = (1.96\ t_{i,N=1}^{\ -1})^2$$

The C_P-efficient design strategy minimises maximum N_i for the environmental attributes of interest.

The benefits of design updating were assessed using a two stage choice experiment undertaken for the purpose of estimating the value of changes in environmental attributes dependent on introduced wasp (*Vespula germanica, V. vulgaris*) management at Lake Rotoiti on the South Island of New Zealand (Kerr and Sharp, 2008).

The Lake Rotoiti area is subject to high wasp populations that thrive in the beech forest particularly because of the prevalence of honeydew (*Ultracoelostoma spp.*), which is an important source of carbohydrate for wasps. Wasps affect recreational experiences because of their aggressive behaviour — often stinging recreationists. Peak wasp biomass is highly significant in these forests and can exceed the combined biomass of birds, rodents and mustelids (Thomas *et al*., 1990). Native wildlife populations are adversely affected by wasps competing for both carbohydrate and protein food sources and because of direct predation by wasps (Beggs, 2001; Beggs and Rees, 1999; Beggs and Wilson, 1991; Harris and Oliver, 1993; Moller, 1990; Toft and Rees, 1998). Biological control and aerial poisoning of introduced wasps has been ineffective to date — the only method available for significantly reducing wasp populations is manual ground application of poison in bait stations, which is both expensive and time-consuming (Beggs *et al*., 1998; Beggs *et al*., 2002; Harris and Rees, 2000).

The benefits of wasp control were investigated using a choice experiment that varied the outcomes of wasp control activities at Lake Rotoiti. Attributes included in the study were the probability of recreationists being stung by wasps on a typical summer or autumn day (5%, 10%, 20%, 50%), the vitality of native bird and insect populations (very low, low, high), and cost. Bird and insect populations were dummy-coded, with low as the base. Cost attribute levels were initially set at $0, $50, $100 and $150, but were changed during the study as more information became available on attribute values. Data were collected in two group meetings held in Christchurch City four nights apart in July 2008. Both groups were drawn from the same population — a local primary school community.

The choice experiment entailed twenty unlabelled choice sets that were presented to all participants. Each choice set consisted of a base alternative (20% probability of being stung, low populations of native birds and native insects, zero cost) and two alternatives to the base. The initial design was developed based on researcher assumptions about WTP developed through focus group and pre-testing procedures. Attribute levels were randomly allocated in a balanced design over the two non-base alternatives. A more efficient design was developed by searching over random rearrangements of the attribute levels, constrained to retain balance. The objective of the search (conducted over 1 million iterations) was to minimise the sample size required to ensure every measure of willingness to pay was significant at the 95% confidence level; C_P-efficiency as defined by Scarpa and Rose (2008). The search process was automated as a macro in Microsoft Excel. While initial design of the process took several days, the search process was relatively rapid, completing about 20,000 iterations per minute on a low specification laptop. This speed allows several million design combinations to be tested in a matter of hours, permitting rapid update for sequential applications. Such speeds are not obtainable with Bayesian updating processes.

In the first stage of data collection the efficient random design was applied to a group of 31 people and a multinomial logit model was estimated for this sample. The second stage of data collection utilised a revised design entailing changes in the cost attribute vector and the experimental design. Second stage data collection used an identical format to the first stage and obtained data from 43 different individuals to those engaged in stage one, but drawn from the same population. In order to remove sample size effects from comparisons of efficiency, sample sizes were equalised by randomly drawing 31 individuals from stage two respondents.

Maddala *et al*. (2003) tested design efficiency by comparison of 95% confidence intervals. A related approach is employed here with the comparison of t-scores for each of the WTP measures estimated at each stage of the survey. Equalisation of sample sizes validates this approach.

The experimental approach entailed drawing two small samples from a large population. Comparison of results from the two samples is therefore potentially confounded by the possibility of underlying taste differences between the two samples. Direct comparison of models derived for the two samples is not possible because of potential scale differences. Two approaches that account for potential scale differences are the "nested logit trick" (Hensher and Bradley, 1993) and the Swait-Louviere test (Swait & Louviere, 1993). The Swait-Louviere test was adopted here using a simple search algorithm to identify optimal scale. This approach entailed creating a new set of attribute levels which were simply the original attributes multiplied by a scale factor. For the stage one data the scale parameter was constrained to be unity, whereas for the second stage the scale parameter was variable. The optimal scale parameter was identified by estimating the model over a large range of scale parameters and identifying the scale parameter value which maximised likelihood of the model. The test statistics, which are distributed chi-squared, are:

$$\lambda_A = -2[L_S - (L_1 + L_2)]$$

$\lambda_B = -2[L_P - L_S]$

Where L_1 and L_2, L_S and L_P are log-likelihood scores for MNL models fitted to data sets 1 and 2, the optimally scaled pooled data, and the data pooled with common scale (the Naïvely Pooled model). Degrees of freedom are k-1 for λ_A, where k is the number of parameters in the MNL model, and one for λ_B. In addition, λ_C tests whether naïve pooling is appropriate.

$\lambda_C = -2[L_P - (L_1 + L_2)]$

3. Results

C-efficiency estimates, using the naïvely pooled model as the base, are reported in Table 1. Using the analysts' priors it was expected that the initial random design would have required a sample size of 38 respondents to estimate each WTP measure with better than 95% confidence of being significantly different from zero. Application of the search algorithm to improve this design resulted in an expected sample size (N=24) of only 63% of the original random sample in order to obtain WTP measures for all attributes significant at the target level. This sample size proved to be overly pessimistic when evaluated against the MNL model coefficients estimated after stage one data collection, which indicated that a sample size of 21 respondents would suffice.

Table 1: Design parameters

Design	Source of priors	Applied	Evaluation	Evaluated against	N	C-Efficiency
Random	Analyst expectations	Not applied	*a priori*	Priors	37.78	24%
Efficient	Analyst expectations	Stage 1	*a priori*	Priors	23.82	38%
Efficient	Analyst expectations	Stage 1	*ex post*	Stage 1 MNL	20.96	43%
Efficient	Stage 1 MNL	Stage 2	*a priori*	Stage 1 MNL	13.72	65%
Efficient	Stage 1 MNL	Stage 2	*ex post*	Stage 2 MNL	11.07	81%
Efficient	Stage 1 MNL	Stage 2	*ex post*	Pooled MNL	11.19	80%
Efficient	Pooled MNL	Not applied	*a priori*	Priors	8.95	100%

The second stage design was enhanced by changes in cost attribute levels. The near absence of native birds was valued more highly than prior expectations, resulting in WTP estimates outside the data range. This result suggested potential benefits from extending the upper limit of the cost attribute. Design investigation entailed use of several different cost attribute vectors and the first stage multinomial logit model coefficient estimates. The result was adoption of a revised cost attribute vector ($0, $50, $150, $250) and a revised experimental design. Expectations were for a 53% increase in C-efficiency[1] over the first stage experimental design (Table 1), reducing the expected sample size to 14 respondents. Again, this expectation was overly pessimistic - a sample of

[1] = 100*[(20.96/13.72)-1]

11 would have attained the stated objective. The potential for further efficiency gains is highlighted by the final row in Table 1, which uses the pooled coefficient estimates as priors and predicts a possible further 25% gain in efficiency. Estimated MNL utility functions are reported in Table 2.

Table 2: MNL models, Christchurch

	Assumed	Stage 1	Stage 2	Naïvely Pooled	Scaled Pooled
Constant	0.15	-0.108	-0.186	-0.116	-0.140
Stings	-0.01	-0.0496***	-0.0519***	-0.0501***	-0.0530***
Very Low Birds	-1.50	-2.082***	-1.698***	-1.920***	-2.044***
High Birds	1.00	1.073***	0.835***	0.947***	1.012***
Very Low Insects	-0.50	-1.046***	-0.901***	-0.936***	-1.019***
High Insects	0.50	0.567***	0.665***	0.641***	0.668***
Cost	-0.01	-0.00678***	-0.00679***	-0.00671***	-0.00716***
Stage 2 scale					.876
N		31	31	62	62
-LL (restricted)		632.570	659.553	1296.854	1296.854
-LL (unrestricted)		478.392	523.791	1005.654	1004.714
McFadden's R^2		.244	.206	.225	.225

* α < .10, ** α < .05, *** α < .01

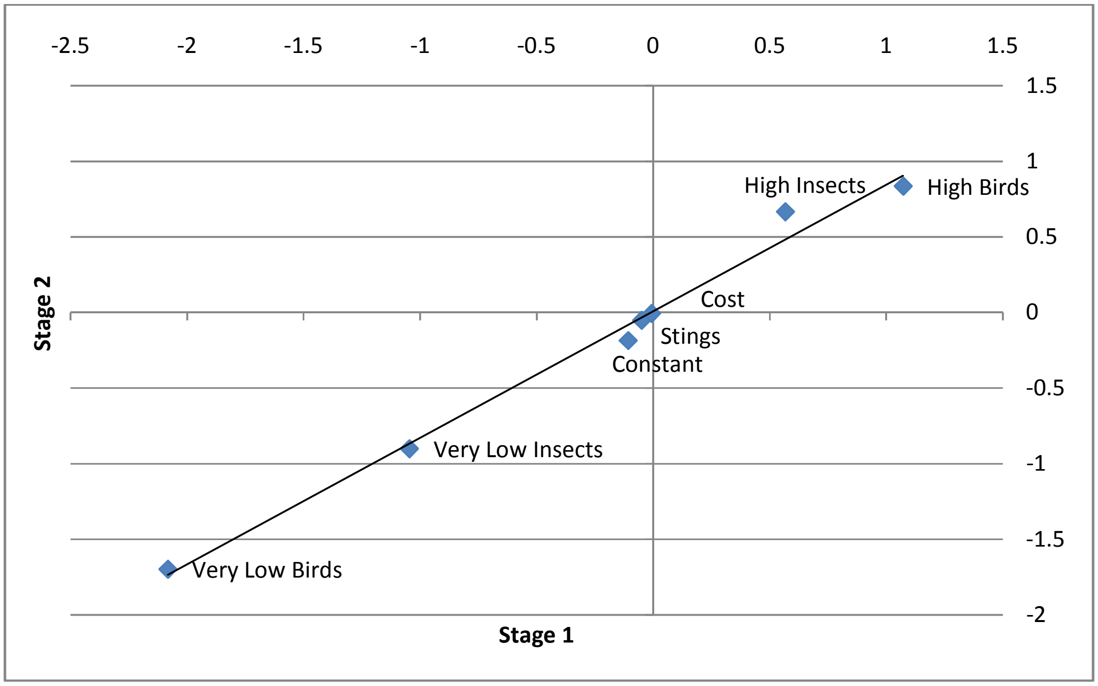
λ_A = 5.062 (6 dof) p=0.536

λ_B = 1.880 (1 dof) p=0.170

λ_C = 6.942 (7 dof) p=0.435

All environmental attribute coefficients are highly significant and of the expected signs. Based on McFadden's R^2, the model estimated for stage two does not fit as well as the model estimated for stage one, although the difference is not great. The Swait-Louviere tests indicate that pooling of the two datasets is appropriate. However, the scale parameter is not significantly different from one and the scaled pooled model does not improve upon the naïvely pooled model. The similarity of the MNL models for stages one and two are further illustrated in Figure 1, which compares utility function coefficients for the two models. Differences in scale preclude direct comparison of these coefficients, but the points will fall on a straight line for identical preference structures (Viney *et al*., 2005). Given uncertainty about the true location of each of the points in Figure 1 there is no reason to suspect that the two survey populations have different values for these environmental attributes.

Figure 1: Comparison of utility function coefficients for stage one and stage two models



The purpose of experimental design updating is to improve estimates of WTP. The tests conducted above indicate that the two samples had the same preferences; consequently, comparison of t-scores and 95% confidence intervals provide valid measures of efficiency. Estimates of mean WTP are presented in Table 3.

Table 3: Mean WTP ($), Christchurch

	Assumed	Stage 1	Stage 2	Naïvely Pooled	Scaled Pooled
Stings	-1	-7.31	-7.65	-7.47	-7.40
Very Low Birds	-150	-307	-250	-286	-286
High Birds	100	158	123	141	141
Very Low Insects	-50	-154	-133	-139	-142
High Insects	50	84	98	95	93

Initial design was undertaken using the WTP values assumed by the researchers (Table 3). Each of the money values assumed by the researchers is less than the corresponding mean WTP measures estimated from survey responses. Consequently, there should be efficiency gains from design updating based on survey data. Mean WTP estimates for stages one and two are not significantly different. Table 4 presents t-scores and standard errors for mean WTP.

Table 4: Absolute t-scores and standard errors for WTP, Christchurch

	Stage 1 N=31	Stage 2 N=31	Improvement at Stage 2
|t-scores|			
Stings	4.11	6.50	58%
Very Low Birds	4.44	5.88	32%
High Birds	4.31	4.57	6%
Very Low Insects	3.65	4.66	28%
High Insects	2.50	3.24	30%
Standard errors			
Stings	1.78	1.18	34%
Very Low Birds	69.14	42.54	38%
High Birds	36.71	26.93	27%
Very Low Insects	42.33	28.49	23%
High Insects	33.51	30.26	10%

The t-scores in Table 4 are all highly significant, even at stage one. It is notable, however, that each of the t-scores improves at stage two, indicative of a more efficient design. The improved t-scores at stage two are reflected in narrower confidence intervals for each WTP measure, ranging from a low of a 10% smaller standard error for high numbers of insects to a 38% reduction for very low bird numbers.

4. Discussion & Conclusions

The sequential data collection employed here led to two improvements in design of the choice experiment. Firstly, the initial application identified the order of magnitude of monetary values associated with the environmental attributes of interest. It became apparent that the cost-attribute vector did not contain sufficiently high values. C_P-efficiency criteria were used to search for the most efficient experimental design across a range of potential cost-attribute vectors. This procedure led to selection of a different cost-attribute vector than was used in stage one, and a new experimental design based on the new cost vector and the initial estimates of utility function coefficients. The substantial improvements in t-scores observed for the stage two multinomial logit model-based estimates of WTP illustrate the benefits of this design updating procedure.

Prior knowledge was used to make assumptions about utility function coefficients. While these estimates were incorrect, they were of the right order of magnitude for four of the five environmental attributes. This relatively close correspondence implies that C-efficiency gains are likely to be relatively minor in this case compared with situations in which prior information is unreliable, or where parameters are assumed to be zero. However, there were still significant gains from redesign, further underlining the potential benefits of the procedure.

Having achieved substantial efficiency gains from a single design update, the question arises as to whether additional updating would be beneficial. That question is easily answered by using coefficient estimates from a pooled model using all of the information obtained to date to optimise the design. The final row of Table 1 indicates that there may be a further efficiency gain in the order of 25% by doing so. If a substantial proportion of the sample remains to be collected such gains would be worth pursuing.

Better prior information reduces the potential gains from sequential design updating. This survey was applied in Nelson City concurrently with second stage data collection in Christchurch. The second stage Christchurch design was used for Nelson City. Nelson values were very similar to Christchurch values. Consequently, improvements in t-scores at stage two were not dramatic, ranging between 4% and 14%, compared with a range of 6% to 58% at Christchurch (Tables 4 and 5). In each case the t-score for high numbers of insects was the lowest in stage one. These increased by 30% and 6% in Christchurch and Nelson respectively. It is notable that, while all t-scores at Nelson improved at stage two, three of five standard errors became larger at stage 2. This result is not inconsistent with the maximisation procedure, which implicitly maximises t-scores, the ratio of WTP and standard error. Consequently, a change in estimated WTP between stages can cause these two measures to move in opposite directions.

Table 5: Absolute t-scores and standard errors for WTP, Nelson

	Stage 1 N=42	Stage 2 N=42	Improvement at Stage 2
|t-scores|			
Stings	8.93	9.25	4%
Very Low Birds	6.63	7.55	14%
High Birds	7.54	7.99	6%
Very Low Insects	8.02	8.43	5%
High Insects	6.07	6.46	6%
Standard errors			
Stings	0.74	0.71	3%
Very Low Birds	65.82	51.46	22%
High Birds	19.49	20.06	-3%
Very Low Insects	25.41	26.40	-4%
High Insects	21.34	21.64	-1%

Observed differences in respondent preferences have led to more widespread use of models that accommodate heterogeneity, including nested logit, latent class and mixed logit models. Bliemer *et al*. (2009) investigated the relationship between model mis-specification and experimental design. Using multinomial logit and nested logit models they showed that designing for one type of model could lead to efficiency losses when another type of model was estimated. The optimisation of designs that assume respondent homogeneity may lead to reduced efficiency of latent class models as the design that caters for the non-existent “typical respondent” becomes less relevant for each of the non-typical groups of respondents. In order to test this potential effect asymptotic t-scores were estimated for two-category latent class models using the stage one and stage two data sets. Results are inconclusive. The stage two design resulted in improvements in t-scores for 8 out of 10 WTP estimates in the latent class model. One of the t-scores that declined was already very high (t=4.8, declining to t=3.8). However, the other was low (t=1.08, declining to t=0.63 for very low insect numbers). Contrary to priors, and the aggregate result, this class of respondents appears to be either unconcerned about native insect population or averse to insects, with positive (but not significant) WTP for a reduction in insect numbers and negative (but insignificant) WTP for increased insect numbers. This aspect needs more work, but there is no reason why an updating process for latent class, or any other type of model, cannot be undertaken. However, it does highlight the importance of identifying the correct model form *a priori*. That can, of course, happen once initial data have been collected if there are sufficient responses to differentiate between model form.

An important research question arises around the matter of what proportion of the survey budget should be expended on initial sampling. On the one hand, sampling more people early on improves estimates of the coefficient vector, leading to the most efficient design for later application. It also provides information useful in determining the correct type of model to estimate – multinomial logit, nested logit, latent class or mixed logit. On the other hand, sampling fewer people initially permits more respondents to complete the updated design, allowing more opportunity to capitalise upon the benefits of improved experimental design. We leave this matter for later scrutiny.

In conclusion, using prior information to improve experimental design is a relatively straightforward and inexpensive task. The advantages expounded in earlier theoretical studies were tested in a field application and were found to yield significant benefits. We commend sequential design updating as a method suitable for alleviating the substantial data collection costs associated with choice experiments, particularly if there is little prior information on parameter values. We encourage further experimental applications of the process, but suggest the need for further research to determine the optimal split of sampling between different stages in data collection and to determine the optimal number of experimental design updates.

Acknowledgements

Funding for this research was provided by the New Zealand Foundation for Science Research and Technology through Nimmo-Bell Limited (PROJ-103230ECOS-NIMMO). The statistical design procedure benefitted greatly from the advice of John Rose, Riccardo Scarpa and Sharon Long.

References

Beggs, J.R. (2001). The ecological consequences of social wasps (*Vespula spp.*) invading an ecosystem that has an abundant carbohydrate resource. *Biological Conservation* 99: 17-28.

Beggs, J.R., Toft, R.J., Malham, J.P., Rees, J.S., Tilley, J.A.V., Moller, H. and Alspach, P. (1998). The difficulty of reducing introduced wasp (*Vespula vulgaris*) populations for conservation gains. *New Zealand Journal of Ecology* 22(1): 55-63.

Beggs, J.R. and Rees, J.S. (1999). Restructuring of Lepidoptera communities by introduced *Vespula* wasps in a New Zealand beech forest. *Oecologia* 119: 565-571.

Beggs, J.R., Rees, J.S. and Harris, R.J. (2002). No evidence for establishment of the wasp parasitoid *Sphecophaga vesparum burra* (Cresson) (Hymenoptera: Ichneumonidae) at two sites in New Zealand. *New Zealand Journal of Zoology* 29: 205-211.

Beggs, J.R. and Wilson, P.R. (1991). The Kaka *Nestor meridionalis*, a New Zealand parrot endangered by introduced wasps and mammals. *Biological Conservation* 56: 23-38.

Bliemer, M.C.J., Rose, J.M. and Hensher, D.A. (2009). Efficient stated choice experiments for estimating nested logit models. *Transportation Research Part B* 43: 19-35.

Ferrini, S. and Scarpa, R. (2007). Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. *Journal of Environmental Economics and Management* 53: 342-363.

Harris, R.J. and Oliver, E.H. (1993) Prey diets and population densities of the wasps *Vespula vulgaris* and *V. germanica* in scrubland-pasture. *New Zealand Journal of Ecology* 17(1): 5-12.

Harris, R.J. and Rees, J.S. (2000). Aerial poisoning of wasps. *Science for Conservation 162*. Department of Conservation, Wellington.

Hensher, D.A. and Bradley, M. (1993). Using stated response data to enrich revealed preference discrete choice models. *Marketing Letters* 4(2): 139-152.

Kanninen, B.J. (1993). Optimal experimental design for double-bounded dichotomous choice contingent valuation. *Land Economics* 69(2): 138-146.

Kerr, G.N. and Sharp, B.M.H. (2008). *Biodiversity Management: Lake Rotoiti Choice Modelling Study*. Agribusiness and Economics Research Unit Research Report No.310, Lincoln University.

Kessels, R., Goos, P. and Vandebroeck, R. (2006). A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research* 43, 409-419.

Louviere, J.J., Hensher, D.A. and Swait, J.D. (2000). Stated choice methods: analysis and application. Cambridge University Press: Cambridge, U.K.

Maddala, T., Phillips, K.A. and Johnson, F.R. (2003). An experiment on simplifying conjoint analysis designs for measuring preferences. *Health Economics* 12: 1035-1047.

Moller, H. (1990). Wasps kill nestling birds. *Notornis* 37: 76-77

Scarpa, R. and Rose, J. (2008). Design efficiency for choice modelling. *Australian Journal of Agricultural and Resource Economics* 52(3), 253-282.

Swait, J. and Louviere, J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research* 30: 305-314.

Thomas, C.D., Moller, H., Plunkett, G.M. and Harris, R.J. (1990). The prevalence of introduced Vespula vulgaris wasps in a New Zealand beech forest community. *New Zealand Journal of Ecology* 13:63-72.

Toft, R.J. and Rees, J.S. (1998). Reducing predation of orb-web spiders by controlling common wasps (*Vespula vulgaris*) in a New Zealand beech forest. *Ecological Entomology* 23; 90-95.

Viney, R., Savage, E. and Louviere, J. (2005). Empirical investigation of experimental design properties of discrete choice experiments in health care. *Health Economics* 14: 349-362.