New Zealand Agricultural and Resource Economics Society (Inc.)

Efficient design for willingness to pay in choice experiments: evidence from the field

Geoffrey N. Kerr
Lincoln University
e-mail: geoffrey.kerr@lincoln.ac.nz

Basil M.H. Sharp
University of Auckland
e-mail: b.sharp@auckland.ac.nz

Paper presented at the 2009 NZARES Conference
Tahuna Conference Centre – Nelson, New Zealand. August 27-28, 2009.

New Zealand Agricultural and Resource Economics Society Conference 2009

Efficient design for willingness to pay in choice experiments: evidence from the field

Geoffrey N. Kerr, Lincoln University

geoffrey.kerr@lincoln.ac.nz

Basil M.H. Sharp, University of Auckland

b.sharp@auckland.ac.nz

Efficient design for willingness to pay in choice experiments: evidence from the field

Geoffrey N. Kerr and Basil M.H.Sharp

Summary

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

Combinations of attributes at different levels create sets of alternatives within a choice experiment. The construction of each alternative and the combinations of alternatives in each choice event is the experimental design. Inappropriate experimental designs may result in unidentifiable choice 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 has been the most common approach to measuring efficiency of experimental designs (Ferrini and Scarpa, 2007). D-efficient designs minimise the D-error, which is an aggregate measure constructed from the variances and covariances of the estimated utility function parameters. Scarpa and Rose (2008) define the D-error as:

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

Ω is the asymptotic variance-covariance matrix for the design variables ($\mathbf{x}_{sj}$), with utility function coefficient vector $\boldsymbol{\beta}$, where s indexes the alternative and j indexes the choice task. K is the number of coefficients estimated. Identification of a D-efficient design entails selection of $\mathbf{x}_{sj}$ which minimises the 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).

Often choice experiments are conducted to identify behaviours or to estimate willingness to pay (WTP). The importance of choice predictions, rather than the utility function *per se*, has been recognised by Kessels *et al*. (2006) who proposed G-optimality and V-optimality based on minimisation of maximum and average choice prediction variances. Similarly, Kanninen (1993) developed designs to minimise the variance in WTP estimates in contingent valuation studies. Recently, Scarpa and Rose (2008) have used design strategies to minimise variance in WTP (C-efficiency) in hypothetical choice experiment simulations which illustrated the potential advantages of designing for C-efficiency, rather than approaches based on D-efficiency and other efficiency criteria.

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 sequential data collection approach uses information obtained in early applications to update the experimental design using either coefficient vector point estimates or by Bayesian updating to account for uncertainty in the coefficient vector. This study empirically estimates C-efficiency gains based on point estimates of the coefficient vector (C_P efficiency, Scarpa and Rose, 2008). The next section describes the methods used. Results are presented in section three. The paper concludes with a discussion of the results and suggestions for further research.

2. Methods

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. If α and β are utility function coefficients for attribute i and cost respectively, mean WTP for attribute i is:

$$WTP_i = -\alpha/\beta$$

Following Scarpa and Rose (2008) the variance of mean WTP may be estimated as:

$$\mathrm{Var}(WTP_i) \approx [\mathrm{Var}(\alpha) + \alpha^2/\beta^2\, \mathrm{Var}(\beta) - 2\, \alpha/\beta\, \mathrm{Cov}(\alpha,\beta)]/\beta^2$$

Where Var(α), Var(β) and Cov(α,β) are elements of the variance-covariance matrix for one replicate of the experimental design. The sample size necessary for mean WTP_i to be significantly different from zero at the γ% significance level is then:

$$N_i = t_{\gamma/2}^2 \, Var(WTP_i)/WTP_i^2$$
$$= t_{\gamma/2}^2[Var(\alpha)/\alpha^2 + Var(\beta)/\beta^2 - 2Cov(\alpha,\beta)/(\alpha\beta)]$$

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

This study assessed the benefits of design updating using a two stage choice experiment undertaken for estimation of values of environmental attributes dependent on introduced wasp (*Vespula germanica, V. vulgaris*) management at Lake Rotoiti (Kerr and Sharp, 2008). The Lake Rotoiti area is subject to high wasp populations that thrive in the beech forest because of the prevalence of honeydew (*Ultracoelostoma spp.*), which is an important source of carbohydrate for wasps. Wasps affect recreational experiences and native wildlife. Peak wasp biomass is highly significant, sometimes exceeding the combined biomass of birds, rodents and mustelids (Thomas *et al*., 1990). 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 Lake Rotoiti wasp control were investigated using a choice experiment that varied the outcomes of wasp control activities. 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 abundance 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 and in meetings held on two consecutive nights in Nelson. In each location, both groups were drawn from the same local primary school community population. The first Nelson application occurred concurrently with the second Christchurch application and utilised the stage 2 Christchurch experimental design.

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 estimate of mean WTP was significant at the 5% level.

In the first stage of data collection the efficient random design was applied to groups of 31 people (Christchurch) and 49 people (Nelson). A multinomial logit model was estimated for these first groups. The second stage of data collection utilised a revised design entailing changes in the experimental design. The cost attribute vector was changed between the two Christchurch stages, but was unaltered at Nelson. Second stage data collection used an identical format to the first stage and obtained data from 43 (Christchurch) and 42 (Nelson) different individuals to those engaged in stage one, but drawn from the same population. Maddala *et al*. (2003) tested design efficiency by comparison of 95% confidence intervals. A related approach is employed here with the comparison of standard errors for each of the mean WTP estimates at each stage of the survey. In order to remove sample size effects from comparisons of efficiency, sample sizes were equalised by randomly drawing individuals from respondents in the stage with the most participants.

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. The Swait-Louviere test (Swait & Louviere, 1993) was used to identify optimal relative scale and to test for differences in preferences for the two samples.

3. Results

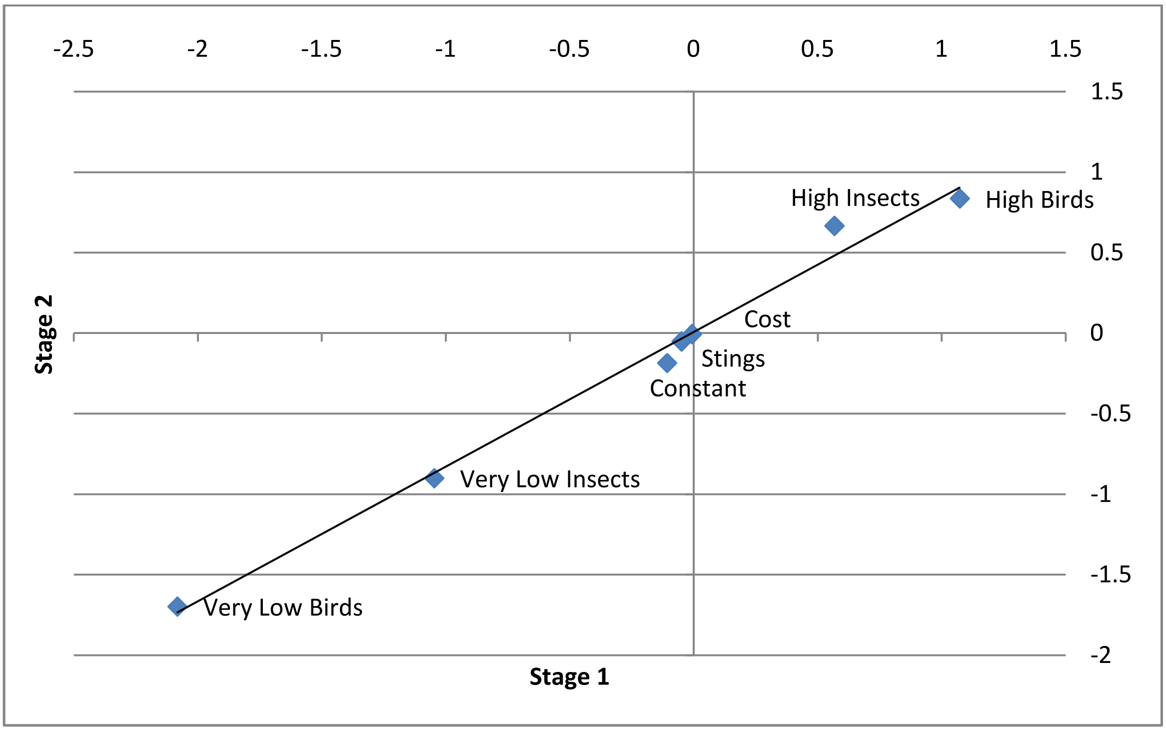
Estimated MNL models for Christchurch are reported in Table 1. All environmental attribute coefficients are highly significant and of the expected signs. The Swait-Louviere test indicated that pooling of the two datasets is appropriate. The optimal relative scale parameter is not significantly different 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. Potential 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.

Table 1: 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***
V 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 relative 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

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



The Christchurch design strategy is reported in Table 2. 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 (Table 2). 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 have sufficed.

Table 2: Design parameters, Christchurch

Design	Source of priors	Applied	Evaluation	Evaluated against	N	C-Efficiency
Random	Analysts	No	*a priori*	Priors	37.78	24%
Efficient	Analysts	Stage 1	*a priori*	Priors	23.82	38%
Efficient	Analysts	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	No	*a priori*	Priors	8.95	100%

The second stage Christchurch 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 (Table 3). 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 2), reducing the expected sample size to 14 respondents. Again, this expectation was overly pessimistic - a sample of 11 would have attained the stated objective. The potential for further efficiency gains is highlighted by the final row in Table 2, which uses the naïvely pooled coefficient estimates as priors and predicts a possible further 25% gain in efficiency.

The tests conducted above indicate that the two samples had similar preferences; consequently, comparison of standard errors provides valid measures of efficiency. Estimates of mean WTP and standard errors are presented in Table 3.

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

Table 3: Mean WTP (Standard error)

	Assumed WTP $	Stage 1 Christchurch $	Stage 2 Christchurch $	Stage 1 Nelson $	Stage 2 Nelson $
Stings	-1	-7.31 (1.78)	-7.65 (1.18)	-6.58 (0.74)	-6.60 (0.71)
Very Low Birds	-150	-307 (69)	-250 (43)	-436 (66)	-389 (51)
High Birds	100	158 (37)	123 (27)	147 (20)	160 (20)
Very Low Insects	-50	-154 (42)	-133 (28)	-204 (25)	-223 (26)
High Insects	50	84 (34)	98 (30)	130 (21)	140 (22)

Experimental design for the first application of the choice experiment 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. It is notable that each of the standard errors for Christchurch improves at stage two, ranging from a 10% smaller standard error for high numbers of insects to a 38% reduction in standard error for very low bird numbers. These results are indicative of a more efficient design and narrower confidence intervals for each estimate of mean WTP. Efficiency gains at Nelson are minor, with only the standard error on "Very Low Birds" improving at stage 2.

4. Discussion & Conclusions

The sequential data collection employed here led to two improvements in design of the choice experiment. Firstly, the initial Christchurch 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-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 revised cost-attribute vector, and a new experimental design based on the new cost vector and the stage one estimates of utility function coefficients. The substantial improvements in standard errors observed for the stage two estimates of WTP illustrate the benefits of this design updating procedure.

Prior knowledge was used to make assumptions about WTP and the related utility function coefficients. While these estimates were incorrect, each dollar value prior being too small, their 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 non-existent. 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 for Christchurch 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 stage one at Nelson City. Because Nelson WTP values were very similar to Christchurch WTP values, improvements at stage two in Nelson were not dramatic. Only one standard error decreased significantly, the others remaining unchanged. It is notable that while standard errors changed little, changes in estimates of mean WTP resulted in improved t-scores for all Nelson WTP estimates, ranging between 4% and 14%. In each case the lowest t-score for stage one was for high numbers of insects. The t-scores for the WTP estimate for this attribute increased by 30% and 6% in Christchurch and Nelson respectively.

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. 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, particularly now that commercial software (Ngene) is now available for the 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 reducing 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

We are extremely grateful to John Rose whose software provided the foundation from which our own algorithm was developed with the assistance of Sharon Long. Riccardo Scarpa provided invaluable conceptual advice and helpful comments that have greatly improved the finished product. Errors remain ours. Funding for this research was provided by the New Zealand Foundation for Science Research and Technology through Nimmo-Bell Limited (PROJ-103230ECOS-NIMMO).

References

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., 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.

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

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.

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.

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.