The Australian Journal of
Agricultural and
Resource Economics

Journal of the Australian Agricultural and Resource Economics Society

The Australian Journal of Agricultural and Resource Economics, 54, pp. 407–420

Choice experiment adaptive design benefits: a case study*

Geoffrey N. Kerr and Basil M. H. Sharp†

Efficient experimental designs offer the potential to reduce required sample sizes, or to reduce confidence intervals for parameters of interest, in choice experiments. Choice experiment designs have typically addressed efficiency of utility function parameter estimates. The recently developed concept of *C*-efficiency recognises the salience of willingness to pay estimates rather than utility function parameters in studies that seek to put money values on attributes. *C*-efficiency design benefits have been illustrated in a theoretical context, but have not been tested in applied settings. This study reports a choice experiment field application that used initial responses to update statistical designs to maximise *C*-efficiency. Consistent with theoretical predictions, the revised design delivered significant reductions in the variance of willingness to pay estimates, illustrating that *C*-efficient designs can indeed decrease costs of choice experiments by reducing required sample sizes.

Key words: *C*-efficiency, choice experiment, experimental design, New Zealand, wasps.

1. Introduction

Choice experiments are a stated preference approach that is in common use for estimating changes in welfare arising from changes in the flow of environmental goods and services. Choice experiments open up the ability to characterise changes in attributes that are likely to follow from a range of projects or policy options. The characteristics of choice alternatives, and how they are combined to make choice sets, identify the experimental design. Experimental design plays an important role in choice experiments because inappropriate designs may result in unidentifiable models or produce biased parameter estimates (Louviere *et al*. 2000; Ferrini and Scarpa 2007; Vermeulen *et al*. 2008). Inefficient experimental designs fail to capture the full extent of information from survey participants, resulting in parameter estimate variances that are larger than potentially achievable with any particular sample size.

* John Rose's Excel software for assessing and improving *D*-efficient designs provided the framework that was modified to address *C*-efficient designs. We thank John for generously sharing his software. Riccardo Scarpa provided inspiration for the research and much helpful advice. Two anonymous referees provided comments that resulted in significant improvements to the original manuscript.

† Geoffrey N. Kerr (email: geoffrey.kerr@lincoln.ac.nz) is at the Department of Environmental Management, Lincoln University, Canterbury, New Zealand. Basil Sharp is at the Department of Economics, University of Auckland, Auckland, New Zealand.

doi: 10.1111/j.1467-8489.2010.00507.x

Choice experiment designs have been investigated using the multinomial logit model, and the nested logit and mixed logit models (Bliemer and Rose 2008; Bliemer *et al.* 2009). There are various design strategies for choice experiments (Carlsson and Martinsson 2003; Street *et al.* 2005; Johnson *et al.* 2007; Rose *et al.* 2008), and choice experiment efficiency criteria are diverse (Scarpa and Rose 2008), with *D*-efficiency the most common measure (Ferrini and Scarpa 2007). *D*-efficient designs minimise the *D*-error, which is defined as:

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

where Ω is the asymptotic variance covariance matrix for the design variables ($\mathbf{x}_{sj}$) for alternatives in choice task *j*. $\boldsymbol{\beta}$ is a vector of K utility function coefficients. Identification of a *D*-efficient design entails arranging the pre-selected elements of $\mathbf{x}_{sj}$ to minimise the *D*-error for some expected $\boldsymbol{\beta}$.

There are several forms of *D*-efficiency, depending upon assumptions about priors on $\boldsymbol{\beta}$. When there is no prior information, $\boldsymbol{\beta}$ is assumed to be zero, leading to D_0 error (also known as D_Z error). Incorporation of non-zero point prior estimates of $\boldsymbol{\beta}$ results in D_P error, whereas incorporation of uncertainty about priors through Bayesian techniques results in D_B error (Ferrini and Scarpa 2007). Huber and Zwerina (1996) illustrated the efficiency costs of failure to utilise information on priors. Information about signs and magnitudes of elements in $\boldsymbol{\beta}$ in non-market valuation studies can come from theory, from information obtained from stakeholders during study design and pre-testing, or from sequential data collection. The latter approach uses information obtained in early applications of a choice experiment to sequentially update the experimental design.

A-efficiency is an alternative to *D*-efficiency that minimises the trace of the asymptotic variance covariance matrix. *A*-efficient designs minimise aggregate parameter variances, but may produce very large covariances (Kessels *et al.* 2006; Scarpa and Rose 2008). *S*-optimal designs minimise the sample size needed for all parameters to be statistically significant (Rose and Bliemer 2005). Kessels *et al.* (2006) have proposed *G*-optimality and *V*-optimality based on minimisation of maximum and average choice prediction variances.

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. Vermeulen *et al.* (2008) estimate what they term as '*WTP-error*', which is a measure of aggregated variance across all estimated measures of WTP. A limitation of *WTP*-error is that WTP variance for individual attributes can be large, despite having minimised aggregate variance, so individual attribute WTP estimates can be non-significant. To address this problem, Scarpa and Rose (2008) applied choice experiment design strategies to minimise variance in WTP for individual attributes (*C*-efficiency). They used hypothetical simulations to illustrate the advantages of designing choice experiments to maximise *C*-efficiency using synthetic data for which true parameter values were known. As with other efficiency measures, *C*-efficiency may be based on either point estimates of the coefficient vector (C_P) or a Bayesian approach (C_B). Efficiency criteria have been evolving to meet diverse researcher needs, which may be related to better understanding of the utility function, or to estimating WTP. Some efficiency measures address aggregate error, whereas others address errors of specific items of interest. The *C*-efficiency measure utilised in this study maximises the worst estimated *t*-score of WTP estimates for the full set of attributes.

Sequential updating of survey design based on information collected in early applications has been suggested as a method for improving efficiency (Raghavarao and Wiley 2006; Ferrini and Scarpa 2007). Scarpa *et al*. (2007) used a sequential Bayesian design updating method in pursuit of *D*-efficient designs. Their results show a 69 per cent reduction in the *D*-error and an average increase of 2.22 in asymptotic *z*-scores for utility function coefficients, underlining the potential benefits of the method.

The theoretical bases for improving the design efficiency of choice experiments have advanced to a stage where the challenge remains to apply the criteria to real-world case studies. To our knowledge, this study is the first of its kind to implement the concept of *C*-optimality in a field setting. This paper focuses on a choice experiment design strategy that minimises the variance of WTP estimates by applying a sequential approach to improving C_P-efficiency by updating the choice experiment design during data collection. The approach is applied in a field setting, in which true parameters are unknown. The next section describes the case study. Section 3 describes theory, data collection and statistical analyses for the sequential design process. Results are presented in Section 4. The paper concludes with a discussion of the results and suggestions for further research.

2. Case study

The benefits of C_P-efficient sequential design updating were assessed using a two-stage choice experiment undertaken to estimate the value of changes in environmental attributes affected by introduced vespid wasp (*Vespula germanica, V. vulgaris*) management at Lake Rotoiti on the South Island of New Zealand (Kerr and Sharp 2008a).

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. 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). Biological control and aerial poisoning of introduced wasps have 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, limiting its applicability (Beggs *et al*. 1998; Harris and Rees 2000). Alternative configurations of wasp control activities produce different benefit flows, motivating the need to understand the benefits associated with different outcomes.

3. Method

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

$$\mathrm{WTP}_i = -\gamma\beta^{-1} \tag{2}$$

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

$$\mathrm{Var}(\mathrm{WTP}_i) \approx \beta^{-2}(\mathrm{Var}(\gamma) - 2\gamma\beta^{-1}\mathrm{Cov}(\gamma,\beta) + \gamma^2\beta^{-2}\mathrm{Var}(\beta)) \tag{3}$$

Using one replicate of the experimental design ($N = 1$) to generate an estimate of the asymptotic variance covariance matrix, it is possible to generate the t-score of the WTP estimate for attribute i:

$$t_{i,N=1} = \mathrm{WTP}_i(\mathrm{Var}(\mathrm{WTP}_i))^{-0.5} \tag{4}$$

The sample size (N_i) necessary for mean WTP_i to be significantly different from zero at α per cent significance level is then:

$$N_i = t_{\alpha/2}^2\mathrm{Var}(\mathrm{WTP}_i)\mathrm{WTP}_i^{-2} = \left(\frac{t_{\alpha/2}}{t_{i,N=1}}\right)^2 \tag{5}$$

C-efficiency minimises variance of WTP_i and therefore minimises the sample size necessary for any desired level of precision in the estimate of WTP_i. Design can be undertaken to address one or several attributes of interest. The C_P-efficient design strategy minimises the maximum N_i for all the environmental attributes of interest (Eqn 5). By construction, the C_P-efficient design also maximises the minimum $t_{i,N=1}$ (Eqn 4).

The benefits of wasp control were investigated using a choice experiment that varied the outcomes of wasp control activities at Lake Rotoiti. Consultation with wasp and wildlife management experts, reviews of literature on wasp ecology, wasp management and Lake Rotoiti conservation reports, and two focus groups, one in Auckland and the other in Christchurch, were used to identify attributes and attribute levels. Participants had no prior knowledge of the subject of the study, and the focus group meetings were split into two stages. The first stage gauged what participants knew about wasps. The second stage provided an overview of wasps, their habitat, affects on the environment and people, and options for management and then investigated participants' views on management. Pre-testing of the survey instrument provided complementary information that was used to develop priors about WTP for each attribute.

Attributes included in the study were the probability of recreationists being stung by wasps on a typical summer or autumn day, the vitality of native bird and insect populations, and cost. Cost attribute levels were changed during the study as more information became available on attribute values. Attribute levels are shown in Table 1.

Data were collected in a survey administered during two meetings that were held with a Christchurch City primary school community four nights apart in July 2008. The samples were drawn from the same population, but were not designed to be representative of the broader community. While this sample selection process is suitable for testing differences in design effects, it is not suitable for drawing inferences about community WTP. Information was collected on individual participants to check for differences in the samples between the two stages.

After an introductory presentation that described wasps, their potential to spread to different environments, their impacts, and methods and costs of control, each group was presented with the choice experiment survey. The choice experiment entailed twenty unlabelled choice sets that were presented to all participants. Each choice set consisted of a base alternative and two alternatives to the base. While 20 choice sets are uncommon, Kerr and Sharp (2008b) employed 27 choice sets with more attributes than used here. Caussade *et al.* (2005) found minimum error variance with about nine choice sets and Arentze *et al.* (2003) failed to detect a significant fatigue effect as choice sets increased in size. Open discussion with the participants after they had completed the choice experiment exercise revealed that they remained engaged throughout the experiment. There was no obvious indication of respondent fatigue.

The initial design was developed based on researcher assumptions about WTP, derived through focus group and pre-testing procedures undertaken

Table 1 Attribute levels

Attribute	Levels
Probability of being stung	5%, 10%, <u>20%</u>, 50%
Native bird population*	Very low, <u>Low</u>, High
Native insect population*	Very low, <u>Low</u>, High
Cost (initial)	<u>$0</u>, $50, $100, $150
Cost (later)	<u>$0</u>, $50, $150, $250

The base case is defined by the underlined attribute levels.
*Dummy-coded, with low as the base.

specifically for this study. Attribute levels were randomly allocated in a balanced design over the two non-base alternatives. A more efficient design was then constructed 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 that every measure of WTP was significant at the 95 per cent confidence level. The search process was automated as a macro in Microsoft Excel, available from the authors on request. 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.

In the first stage of data collection, the efficient random design was applied to a sample of 31 people. A multinomial logit model was estimated for this sample. The second stage of data collection utilised information from the first stage to update the design by changing both the cost 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. To remove sample size effects from comparisons of efficiency, sample sizes were equalised by randomly drawing 31 individuals from stage two respondents. To test the efficiency of our survey design, the 95 per cent confidence intervals and the asymptotic *t*-scores for each of the WTP estimates were compared across the various stages of the experiment (Maddala et al. 2003; Scarpa *et al.* 2007). 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 potentially confounded by the possibility of underlying preference differences between the two samples (Viney *et al.* 2005). Direct comparison of models derived for the two samples is not possible because of potential differences in scale and underlying preferences. Scale differences are related to respondent efficiency, with smaller scale parameters indicating more variance in the data and less respondent efficiency (Maddala *et al.* 2003). The Swait-Louviere test (Swait and Louviere 1993) was used to identify the optimal scale ratio for the two data sets. This approach entailed creating a new set of attribute levels that 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 ratio parameter was identified by estimating the model over a large range of scale ratio parameters and identifying the scale ratio parameter value that maximised likelihood of the model.

4. Results

The sample was comprised of 40 per cent men, 76 per cent New Zealand Europeans and 13 per cent Maori, 43 per cent held a university degree, 20 per cent had visited Lake Rotoiti in the previous 5 years, and 8 per cent were members of environmental groups. The mean age was 46 years, and mean personal income was in the range $40 000–50 000 per annum. The only difference between participants in the two stages that was significant at the 10 per cent level was the number of people under the age of 18 years in the household, with means of 1.9 at stage one and 1.2 at stage two ($Z = 1.89$).

Discussions in two focus groups provided initial estimates of WTP for each attribute. These WTP estimates did not exceed $150 for any of the attributes, which was therefore adopted as the upper bound on the original cost vector. The marginal utility of costs was based on analyst assumptions and previous choice experiment experiences. The prior estimate for the cost coefficient was assumed to be −0.01 and the prior estimates for each attribute coefficient were calculated by dividing expected mean WTP_i by 100. Contrary to the expectations from the focus group discussions, results from stage one indicated that one attribute was valued in the order of $300 and two others were valued at about $150. The WTP estimates were therefore outside the data range, suggesting potential benefits from extending the upper limit of the cost vector. C_{P}-efficient designs were created using two candidate cost vectors ($0, $50, $150, $250) and ($0, $100, $200, $300) and the first stage multinomial logit model utility function coefficient estimates. The alternative cost vector ranging from $0 to $250 yielded the lowest sample size and was selected as the cost vector for the second sampling stage. One potential effect of increasing the magnitude of components of the cost vector and changing the experimental design is an increase in protest responses and choice of the base alternative. This was not the case. Only one of the 1240 choices was not answered and that occurred at stage one. Selection of the base case declined from 52 per cent at stage one to 47 per cent at stage two, a change that is significant at the 5 per cent level.

Table 2 shows the sample size effects of design enhancements throughout the study. *C*-efficiency estimates are the ratio of sample sizes for the most

Table 2 Design parameters: Christchurch

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	12.19	73
Efficient†	Stage 1 MNL	Stage 2	*a priori*	Stage 1 MNL	10.90	82
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

*Original cost vector.
†Revised cost vector.

efficient design (the naïvely pooled model) and each other design. Using the analysts' priors, the initial random design would have required a sample size of 38 respondents to estimate each WTP measure at 95 per cent confidence. Application of the search algorithm to improve this design resulted in an expected sample size ($N = 24$) of only 63 per cent of the original random sample to obtain WTP measures for all attributes significant at the target level. This sample size prediction proved to be realistic when evaluated against the multinomial logit model coefficients estimated after stage one data collection, which indicated that a sample size of 21 respondents would have sufficed.

Expectations for stage two, utilising the new cost vector and new experimental design were for a sample size of only 52 per cent of that required using the first-stage experimental design (Table 2), reducing the expected sample size to 11 respondents. The information obtained at stages one and two can be used to further refine experimental design. In the final row in Table 2, it is shown that using prior estimates from the combined stage one and stage two sampling may lead to further gains in efficiency of about 25 per cent if a third sampling round was to be implemented.

Estimated multinomial logit utility functions are reported in Table 3. All environmental attribute utility function coefficients are highly significant and of expected signs. The pooled model with the optimal scale ratio was compared with two independent models using a likelihood ratio test ($\chi^2 = 5.06$, 6 d.f., $P = 0.54$). The optimal scale ratio parameter is not significantly different from stage one, and the scaled pooled model does not improve significantly upon the naïvely pooled model ($\chi^2 = 1.88$, 1 d.f., $P = 0.17$). The Swait-Louviere tests indicate that pooling of the two datasets is appropriate. The similarity of the multinomial logit 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). Figure 1 suggests there is no reason to suspect that the two survey populations have different preferences for these environmental attributes. Together, Figure 1 and the Swait-Louviere tests indicate preferences are similar across samples and there is no diminution of respondent efficiency because of the revised design (Viney *et al*. 2005; Scarpa *et al*. 2007).

The purpose of experimental design updating is to improve efficiency of WTP estimates for a given sample size. Mean WTP estimates for stages one and two are not significantly different (Table 4). The tests conducted earlier indicate that the two samples had the same preferences; consequently, comparison of 95 per cent confidence intervals or standard errors provide valid measures of efficiency of WTP estimates (Scarpa *et al*. 2007).

The improved estimation efficiency at stage two is reflected in the narrower confidence intervals for each WTP estimate, ranging from a 10 per cent smaller standard error for high numbers of insects to a 38 per cent reduction in standard error for very low bird numbers. The design efficiency is further

Table 3 MNL model parameters, Christchurch (standard errors)

	Assumed	Stage 1	Stage 2	Naively pooled	Scaled pooled
Constant	0.15	−0.108 (0.194)	−0.186 (0.210)	−0.116 (0.139)	−0.140 (0.148)
Stings	−0.01	−0.0496*** (0.0051)	−0.0519*** (0.0054)	−0.0501*** (0.0036)	−0.0530*** (0.0038)
Very low birds	−1.50	−2.082*** (0.229)	−1.698*** (0.216)	−1.920*** (0.154)	−2.044*** (0.164)
High birds	1.00	1.073*** (0.184)	0.835*** (0169)	0.947*** (0.121)	1.012*** (0.130)
Very low insects	−0.50	−1.046*** (0.203)	−0.901*** (0.186)	−0.936*** (0.135)	−1.019*** (0.144)
High insects	0.50	0.567*** (0.200)	0.665*** (0.171)	0.641*** (0.125)	0.668*** (0.134)
Cost	−0.01	−0.00678*** (0.00149)	−0.00679*** (0.00103)	−0.00671*** (0.00082)	−0.00716*** (0.00088)
Stage 2 scale					0.876
N		31	31	62	62
-LL (restricted)		632.570	659.553	1296.854	1296.854
-LL (model)		478.392	523.791	1005.654	1004.714
McFadden's R^2		0.244	0.206	0.225	0.225

$*\alpha < 0.10$, $**\alpha < 0.05$, $***\alpha < 0.01$.

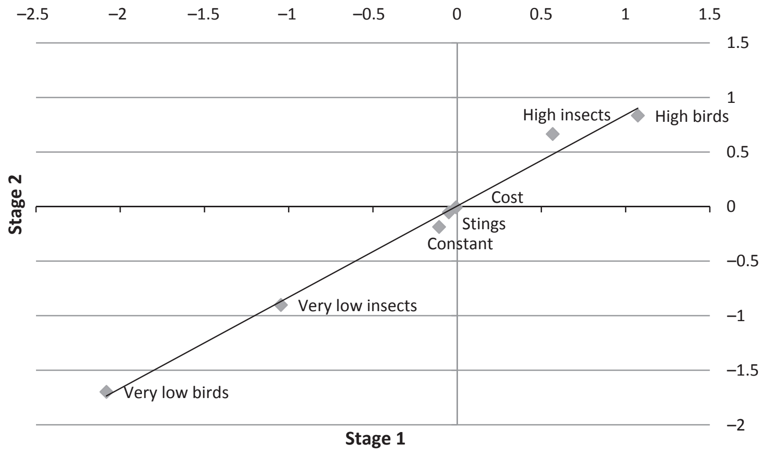

Figure 1 Comparison of utility function coefficients for stage one and stage two models.

Table 4 Mean WTP estimates, Christchurch

	Stage 1	Stage 2	Difference
Stings			
Mean WTP	−$7.31	−$7.65	$0.34
Std error	$1.78	$1.18	$2.13
|*t*-score |	4.11	6.50	0.16
Very low birds			
Mean WTP	−$307	−$250	$57
Std error	$69	$43	$81
|*t*-score|	4.44	5.88	0.70
High birds			
Mean WTP	$158	$123	$35
Std error	$37	$27	$46
|*t*-score |	4.31	4.57	0.77
Very low insects			
Mean WTP	−$154	−$133	$22
Std error	$42	$28	$51
|*t*-score |	3.65	4.66	0.42
High insects			
Mean WTP	$84	$98	$14
Std error	$34	$30	$45
|*t*-score |	2.50	3.24	0.32

demonstrated by the highly significant *t*-scores for all WTP estimates, and the improvements in *t*-scores at stage two of the experiment.

To test the effect of quality of prior information on efficiency improvements from sequential design updating the survey was later applied in Nelson City using the second stage Christchurch design. If mean WTP is similar in both locations, then efficiency gains at Nelson from sequential updating are expected to be less than efficiency gains at Christchurch. The mean WTP estimates in Nelson were similar to Christchurch estimates (Table 5).

Table 5 Mean WTP estimates, Nelson

	Stage 1	Stage 2	Difference
Stings			
Mean WTP	−\$6.58	−\$6.60	\$0.02
Std error	\$0.74	\$0.71	\$1.03
|*t*-score |	8.93	9.25	0.02
Very low birds			
Mean WTP	−\$436	−\$389	\$48
Std error	\$66	\$51	\$84
|*t*-score |	6.63	7.55	0.57
High birds			
Mean WTP	\$147	\$160	\$13
Std error	\$19	\$20	\$28
|*t*-score |	7.54	7.99	0.48
Very low insects			
Mean WTP	−\$204	−\$223	\$19
Std error	\$25	\$26	\$37
|*t*-score |	8.02	8.43	0.51
High insects			
Mean WTP	\$130	\$140	\$10
Std error	\$21	\$22	\$30
|*t*-score |	6.07	6.46	0.34

Improvements in *t*-scores at stage two in Nelson were not dramatic, ranging between 4 per cent and 14 per cent, compared with a range of 6 per cent to 58 per cent at Christchurch (Tables 4 and 5). It is notable that, while all *t*-score s at Nelson improved at stage two, three of five standard errors became larger at stage two. This result is not inconsistent with the *C*-efficient design procedure, which implicitly maximises the *t*-scores of the estimated WTPs. Consequently, an increase in the estimated mean WTP between stages can result in higher *t*-scores as well as larger standard errors. All three Nelson attributes for which standard errors increased at stage two had higher expected mean WTP at stage two. These results are consistent with the hypothesis that better prior information increases the initial *C*-efficiency and reduces potential gains from *C*-efficient sequential design updating.

5. Discussion and conclusions

The sequential data collection employed here led to two improvements in design of the choice experiment. First, the initial application identified the order of magnitude of monetary values associated with the environmental attributes of interest. It became apparent that the cost vector did not contain sufficiently high values. C_P-efficiency criteria were used to search for a more efficient experimental design across a range of potential cost vectors. This procedure led to selection of a different cost vector and a new experimental design that was based on the new cost vector and initial estimates of utility function coefficients. The substantial improvements in *t*-scores and standard errors observed for the stage two multinomial logit model WTP estimates illustrate the benefits of this design updating procedure. Because the cost vector and the experimental design changed concurrently, it is not possible to identify the contribution that each made to the total improvement in efficiency. With the original cost vector, predicted minimum sample size decreased 42 per cent with application of a *C*-efficient design to stage one parameter estimates. Changing the cost vector resulted in a further 11 per cent decrease in minimum sample size. Further opportunities to enhance efficiency may be realised by experimentation with levels for other numeric attributes – in this case the stings attribute. The number of attribute levels and their distribution are both candidates for experimentation in future studies.

Further experimental applications of this process are needed to jointly determine the optimal proportion of the sample that should be allocated to each stage of data collection, and the optimal number of experimental design updates. 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. On the other hand, with a fixed data collection budget, sampling fewer people initially permits more respondents to complete the updated design, allowing more opportunity to capitalise upon the benefits of improved experimental design. Pretesting of survey instruments offers one opportunity to obtain prior information on WTP that can be used for initial design purposes. The small numbers utilised in our study indicate that, even when pretests are small, information obtained from them can be helpful – although representativeness issues need to be borne in mind. Data collection methods will also have some bearing on this matter. For example, an internet-based survey or personal interviews can be cheaply and quickly adapted to incorporate modified experimental designs. In such cases, redesign can be a continuous process at little cost. The relatively high cost of personal interviews and their sequential nature suggest that they will offer the greatest cost and time benefits from minimisation of sample sizes and investment in multiple experimental design updates will be most worthwhile with this data collection mechanism. Paper-based data collection methods, such as drop-off, pick-up and postal surveys, cannot as easily incorporate new designs because of the requirement to print and distribute modified surveys. Paper-based data collection methods may benefit from more extensive piloting to obtain better priors. We leave this matter for later scrutiny.

The best model specification is commonly not known before data collection, but is an important matter to address in experimental design. The efficient design process undertaken here addressed only the multinomial logit model, which may not be the best model for the data. Bliemer *et al.* (2009) compared designs for multinomial and nested logit models, finding that designs for one model did not perform efficiently for the other. Rose *et al.* (2009) came to the same conclusion for multinomial logit, mixed logit and error component models. They addressed uncertainty about correct model specification with a model averaging approach, selecting designs that are robust to model specification. An alternative to model averaging is to use early data, collected using a model averaging design, to identify the correct type of model to estimate, then use initial responses to create an optimal design for the correct model type. Again, the question of how much data to collect at each stage of the process arises. This is an issue for future investigation.

Using prior information to improve experimental design is a relatively straightforward task. Earlier studies on experimental designs have demonstrated the theoretical advantages of updating information. The study presented in this paper found significant benefits from using sequential updating in a *C*-efficient experimental design using data from a field application in New Zealand. 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. The analyst expectations of the prior parameter assumptions were in the same order of magnitude as the survey results for four of the five environmental attributes. However, this study still showed significant gains from survey redesign, underlining the benefits of the sequential updating procedure. The *C*-efficiency gains demonstrated here are likely to be even higher when prior information about parameters is unreliable or unavailable.

References

Arentze, T., Borgers, A., Timmermans, H. and DelMistro, R. (2003). Transport stated choice responses: effects of task complexity, presentation format and literacy, *Transportation Research Part E* 39, 229–244.

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.

Bliemer, M.C.J. and Rose, J.M. (2008). *Construction of Experimental Designs for Mixed Logit Models Allowing for Correlation Across Choice Observations*, Working Paper ITLS-WP-08-13, Institute of Transport and Logistics Studies, University of Sydney, Sydney.

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.

Carlsson, F. and Martinsson, P. (2003). Design techniques for stated preference methods in health economics, *Health Economics* 12, 281–294.

Caussade, S., Ortúzar, J., Rizzi, L.I. and Hensher, D.A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates, *Transportation Research Part B* 39, 621–640.

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.

Huber, J. and Zwerina, K. (1996). The importance of utility balance in efficient choice designs, *Journal of Marketing Research* 39, 214–227.

Johnson, F.R., Kanninen, B.J., Bingham, M. and Osdemir, S. (2007). Experimental design for stated choice studies, in Kanninen, B.J. (ed), *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice*. Springer, Dordrecht, pp. 159–202.

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

Kerr, G.N. and Sharp, B.M.H. (2008b). Evaluating off-site environmental mitigation using choice modelling, *Australian Journal of Agricultural and Resource Economics* 52(4), 381–399.

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, UK.

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.

Raghavarao, D. and Wiley, J.B. (2006). Design strategies for sequential choice experiments involving economic alternatives, *Journal of Statistical Planning and Inference* 136, 3287–3306.

Rose, J.M. and Bliemer, M.C.J. (2005). *Sample Optimality in the Design of Stated Choice Experiments*. Working Paper ITLS-WP-05-13. Institute of Transport and Logistics Studies, University of Sydney, Sydney.

Rose, J.M., Bliemer, M.C.J., Hensher, D.A. and Collins, A.T. (2008). Designing efficient stated choice experiments in the presence of reference alternatives, *Transportation Research Part B* 42, 395–406.

Rose, J.M., Scarpa, R. and Bliemer, M.C.J. (2009). *Incorporating model Uncertainty into the Generation of Efficient Stated Choice Experiments: A Model Averaging Approach*. Working Paper ITLS-WP-09-08. Institute of Transport and Logistics Studies, University of Sydney, Sydney.

Scarpa, R. and Rose, J. (2008). Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why, *Australian Journal of Agricultural and Resource Economics* 52(3), 253–282.

Scarpa, R., Campbell, D. and Hutchinson, W.G. (2007). Benefit estimates for landscape improvements: sequential Bayesian design and respondents' rationality in a choice experiment, *Land Economics* 83(4), 617–634.

Street, D.J., Burgess, l. and Louviere, J.J. (2005). Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments, *International Journal of Research in Marketing* 22, 459–470.

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.

Vermeulen, B., Goos, P., Scarpa, R. and Vandebroek, M. (2008). *Efficient and Robust Willingness-to-Pay Designs for Choice Experiments: Some Evidence from Simulations. KBI0809*. Department of Decision Sciences and Information Management, Katholieke Universiteit Leuven, Leuven, Belgium.

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.