DISCUSSION PAPER

April 2006 ■ RFF DP 05-60

Discrete Choice Survey Experiments

A Comparison Using Flexible Models

Juha Siikamäki and David F. Layton

RESOURCES FOR THE FUTURE

1616 P St. NW
Washington, DC 20036
202-328-5000 www.rff.org

Discrete Choice Survey Experiments: A Comparison Using Flexible Models

Juha Siikamäki and David F. Layton

Abstract

This study investigates the convergent validity of discrete choice contingent valuation (CV) and contingent rating/ranking (CR) methods using flexible econometric methods. Our results suggest that CV and CR can produce consistent data (achieve convergent validity) when respondent's preferred choices and the same changes in environmental quality are considered. We also find that CR models that go beyond modeling the preferred choice and include additional ranks cannot be pooled with the CV models. Accounting for preference heterogeneity via random coefficient models and their flexible structure does not make rejection of the hypothesis of convergent validity less likely, but instead rejects the hypothesis to about the same degree or perhaps more frequently than fixed parameter models commonly used in the literature.

Key Words: valuation, stated preferences, data pooling, random coefficients, Rayleigh, habitat conservation

Contents

Discrete Choice Survey Experiments: A Comparison Using Flexible Models

Juha Siikamäki and David F. Layton*

1. Introduction

Environmental economists use a variety of survey methods for estimating the economic value of environmental quality. Despite differences in format, all survey-based, nonmarket valuation methods attempt to measure the tradeoffs between money and environment, thus estimating willingness-to-pay (WTP) for the changes in environmental quality. The most widely applied nonmarket valuation method is discrete choice contingent valuation, followed by contingent choice, contingent rating, and contingent ranking methods (we will use CV to refer to discrete choice contingent valuation, and CR to refer to contingent ranking and rating). The CV method is based on asking for the acceptance or refusal of a policy program that has specified costs and environmental outcomes. The CR method asks survey respondents to rate or rank alternative policies or, at the simplest, to choose between the status quo and a new program. Since these methods observe stated rather than actual preferences, they also are broadly categorized as stated preference (SP) methods.

If all survey-based valuation methods recover the same underlying preferences, then different SP methods should produce statistically indistinguishable WTP estimates when the same changes in environmental quality are being examined. However, many studies have found that the CR methods estimate higher WTP than the CV method (Desvousges and Smith 1983;

* Juha Siikamäki is at Resources for the Future and can be reached at juha@rff.org. David F. Layton is at the Daniel J. Evans School of Public Affairs, University of Washington, Seattle. We are extremely grateful to Douglas Larson and James Wilen for their help and input throughout this research. The authors would also like to acknowledge helpful comments and suggestions by a coeditor and anonymous referees of the *Journal of Environmental Economics and Management*, which has accepted this paper for publication. Useful comments to earlier versions of this paper by Anni Huhtala and Martin Smith are also greatly appreciated. We would also like to thank the many colleagues in Finland and the United States who reviewed and commented on different questionnaire versions. Siikamäki received funding for this research from Suomen Akatemia (The Academy of Finland) and Yrjö Janssonin Säätiö (Yrjö Jahnsson Foundation). Siikamäki also thankfully acknowledges support by the MTT Economics at Agrifood Research Finland during the survey design and implementation.

Magat et al. 1988; Ready et al. 1995; Stevens et al. 2000). On the other hand, Boxall et al. (1996) found that the CV method sometimes estimates an order of magnitude higher WTP than the CR.[1]

Recently, econometric methods for pooling data from different sources, which have been applied earlier in marketing (Louviere et al. 2000; Hensher et al. 1999) and for combining data on revealed and stated recreational site choices (Adamowicz et al. 1994; Cameron 1992; Kling 1997, Huang et al. 1997; Azevedo et al. 2003; Boxall et al. 2003), have also been used for modeling combined data from different SP surveys (Adamowicz et al. 1998; Cameron et al. 2002; Halvorsen 2000). These studies have demonstrated that SP data from different sources can be modeled using the same underlying econometric model. The data pooling methods are attractive for the purpose of comparing different SP methods since they enable researchers to test for the differences between SP methods within the econometric framework itself, rather than making comparisons between the results of separate econometric models.

While researchers have already highlighted the potential of using data pooling methods in testing for the convergent validity of different SP methods, earlier empirical evidence relies on econometric methods that embody restrictive assumptions and properties. All studies on the topic are based on using fixed, coefficient discrete choice models, typically logit models,[2] and linear functional forms in modeling utility from changes in environmental quality. In addition, except for Stevens et al. (2000) and Cameron et al. (2002), recent evidence on differences between SP methods comes from surveys that ask the same respondents different types of SP questions (e.g., Adamovicz et al. 1998; Halvorsen 2000). Possible concern with data based on with-in subject comparisons is that any consistency between elicitation methods may be an artifact of asking the same people to do both tasks. This concern is avoided by using a split-sample survey design and randomly assigning different survey formats to different respondents, which is the approach we take in this study.

The purpose of this paper is to examine whether CV and CR methods produce data on environmental preferences that are statistically indistinguishable. In order to make comparisons between the CV and CR methods, we designed a split-sample mail survey experiment in which the questionnaires are exactly the same across the whole sample except for the valuation question. In the

---

[1] In addition, Welsh and Poe (1998) develop a multiple, bounded discrete choice CV format that asks respondents to state their preference certainty. In comparing the values obtained by the new format with values from dichotomous choice, payment card, and open-ended CV methods, the authors find that the multiple, bounded format covers the range of values associated with the other three elicitation methods.

[2] The assumptions and properties of fixed coefficient models are restrictive, but more flexible discrete choice models were impractical earlier due to limitations in computing power and simulation-based econometrics (Train 1998, 2001, 2003; Chen and Cosslett 1998; Layton 2000; Layton and Brown 2000).

valuation question, we use either a CV or a CR format but examine preferences for the same and multiple changes in environmental quality. Our empirical analysis uses flexible econometric methods and data pooling models. The flexibility of our econometric methods comes from two sources: First, we apply random coefficient models that relax the restrictions of fixed coefficient models, and second, we consider a variety of non-linear indirect utility specifications that allow for non-linear WTP. We use both an experimental design and econometric methods in such a way that they should have a minimal impact on the rejection or acceptance of the equality between the CV and CR data.

This paper focuses on comparing CV and CR methods in their commonly applied formats. There are several important research topics in both CV and CR literature which we therefore do not seek to explore in this paper. In CV analysis, for example, we exclude starting point effects to estimate exactly similar model specifications for CV and CR data. By doing so, we seek to ensure that differences in the model specifications do not influence our conclusions. Nevertheless, starting point effects are clearly an important research topic (see Herriges and Shogren 1996; Cummings et al. 1997; Haab et al. 1999). In the CR analysis, we refer to earlier studies that have already examined the CR methods. For example, Boyle et al. (2001) used a split-sample design to evaluate the consistency of ratings, rankings, and choice formats in conjoint analysis questions. Layton (2000) used random coefficient models for analyzing SP data, including examining the consistency of rankings. Other studies include Chapman and Staelin (1982) and Hausman and Ruud (1987), which develop estimation methods and test for the consistency of rank-ordered CR data; Ben-Akiva et al. (1991), which investigates the reliability of stated rankings data and finds that the stability of ranking information decreases with decreasing rank; and Foster and Mourato (2002), which tests for inconsistent rankings in conjoint data. These studies investigate the properties of alternative CV and CR approaches but do not relate CR and CV methods, which is the goal of this paper.

Our empirical application deals with measuring WTP for protecting forest habitats that are especially valuable ecologically (biodiversity hotspots) in the non-industrial, private forests of Finland. These areas cover a total of 1.1 million hectares, which is about six percent of Finnish forests. Current regulations and conservation programs protect some 120,000 hectares of hotspots, and extending protection to additional areas currently is being debated. The empirical application evaluates potential conservation policy alternatives for the future by examining the public’s preferences for them.

This paper also introduces a novel way of estimating bounded random coefficients. Although random coefficient models are now rather easily applicable, estimating strictly negative or positive coefficients as lognormals has turned out to be difficult and sometimes impossible in practice (Train 2001; Hensher and Green 2003). These estimation problems limit empirical work, since on theoretical grounds researchers typically want to impose sign restrictions on some parameters; for instance, on those that measure the marginal utility of income. Train (2001) and Train and Sonnier (2005) develop and discuss Bayesian alternative methods for estimating bounded random coefficients. However, in the classical framework used in most empirical applications, complicated random coefficient distributions may not be identifiable in practice. We demonstrate how to use the single-parameter Rayleigh distribution in estimating bounded random coefficients, which can be useful, in particular, if the researcher is hesitant to turn to Bayesian estimation methods.

We are not aware of other comparisons of CV and CR methods that use a split-sample survey design, apply both methods to evaluate exactly the same and multiple changes in environmental quality, and employ random coefficient models. For example, Adamovicz et al. (1998) and Halvorsen (2000) use within-sample comparisons. Cameron et al. (2002) combine data from several CV-type questions regarding one program with CR data on multiple programs, out of which only one program is common with the CV data. The application in this paper elicits preferences for three programs by using both CV and CR. In addition, all past comparisons of stated preference methods have used fixed coefficient models, which is a restriction relaxed in this paper.

Generally speaking, our findings are supported by the results of the Cameron et al. (2002) study, which accepts the convergent validity of several preference elicitation methods based on discrete choices. Our results show that CV and CR methods can produce consistent data and achieve convergent validity when respondent's preferred choices and same changes in environmental quality are considered. But we also find that CR models that go beyond the preferred choice cannot be pooled with the CV models. Accounting for preference heterogeneity via random coefficient models with their more flexible structure does not make rejection of the hypothesis of convergent validity less likely, but instead rejects the hypothesis to about the same degree or perhaps more frequently than fixed parameter models commonly used in the literature.

The rest of the paper is organized as follows. Section 2 explains the econometric models for discrete choice data. Section 3 describes our survey. Section 4 explains the estimation of our models. Section 5 first presents the results of separate CV and CR models and then describes the results of pooled models, including the tests of convergent validity of CV and CR data. Section 6 reports the WTP estimates and Section 7 concludes the paper.

2. Econometric Models For Discrete Choice Data

Econometric models for stated preference survey responses typically are based on McFadden's (1974) random utility model (RUM), which we also use as the point of departure for explaining various econometric models for CV and CR survey responses. We refer readers to Haab and McConnell (2002) for a comprehensive treatment of the econometric methods for modeling stated preference data and describe these methods only to the extent necessary for explaining our empirical analysis.

Consider an individual i choosing a preferred alternative from a set of m alternatives providing utility U_{ij}, that can be additively separated into an unobserved stochastic component ε_{ij} and a deterministic component $V_{ij}(z_j,y_i-A_{ij})$, that is, the indirect utility function that depends only on individual's income y and environmental quality z. Denoting the cost of alternative j to person i with A_{ij}, the utility of alternative j can then be represented as $U_{ij} = V_{ij}(z_j,y_i-A_{ij}) + \varepsilon_{ij}$, where the stochastic term ε_{ij} represents the unobserved factors affecting the choices.

Choices are based on utility comparisons between the available alternatives, and the alternative providing the highest utility becomes the preferred choice. The probability of person i choosing alternative j from among the m alternatives therefore equals the probability that alternative j provides person i with a greater utility U_{ij} than any other available alternative U_{ik}:

$$P_{ij} = P(U_{ij} > U_{ik},\ k = 1, ..., m, \forall\ k \neq j). \tag{1}$$

2.1. Fixed Coefficient Models

Assuming that ε_{ij} and ε_{ik} are independently and identically distributed (i.i.d.), type I extreme value random variables yield the familiar conditional logit model (McFadden 1974) with choice probabilities:

$$P_{ij} = \frac{e^{\mu X_{ij}\beta}}{\sum_{k=1}^{k=m} e^{\mu X_{ik}\beta}}. \tag{2}$$

Parameter μ in (2) is a scale factor, which is implicit in choice models based on the RUM. With data from a single source, μ typically is normalized to one and the normalized parameter vector β is estimated (without the restriction imposed on μ, neither μ nor β could be identified).

Beggs et al. (1981) and Chapman and Staelin (1982) extended (2) to modeling the ranking of alternatives. A rank-ordered logit model treats the ranking as *m-1* consecutive conditional choice problems. It assumes that the ranking results from *m-1* utility comparisons, where the highest ranking is given to the best alternative (the preferred choice from the available alternatives), the second highest ranking to the best alternative from the remaining *m-1* alternatives, the third ranking to the next best alternative, and so on. Indexing the *m* alternatives from best to worst, the probability of ranking *r* by person *i* is given by:

$$P_{ir} = \prod_{j=1}^{m-1} \frac{e^{\mu X_{ij}\beta}}{\sum_{k=j}^{k=m} e^{\mu X_{ik}\beta}} \tag{3}$$

The standard conditional logit formula (2) also applies for modeling CV data, since from the point of view of RUM, a CV response results from a utility comparison between two alternatives: status quo and a new policy.[3] Using a single-bounded (SB) CV and status quo as the reference utility level normalized to zero, conditional logit probabilities of *Yes*- and *No*-answers are $P_i(Yes) = e^{\mu X_{Bid}} \big/ (e^{\mu X_{Bid}} + 1)$ and $P_i(No) = 1 - P_i(Yes)$, where X_{BID} denotes the attributes of the new policy. When CV is applied in the double-bounded (DB) format, respondents are asked a follow-up question based on the first response. Persons who answer *Yes* to the first question (*FirstBid)* are asked a second WTP question with *HighBid* > *FirstBid* and persons who answer *No* get a second question with *LowBid* < *FirstBid*. The four possible response sequences are therefore *Yes-Yes*, *Yes-No*, *No-Yes* and *No-No*. Denoting the exogenous variables for questions with *FirstBid*, *HighBid* and *LowBid* by X_{iFB}, X_{iHB} and X_{iLB}, the possible response probabilities are $P_i(YY) = e^{\mu X_{iHB}\beta} / (e^{\mu X_{iHB}\beta} + 1)$, $P_i(YN) = (1 + e^{\mu X_{iHB}\beta})^{-1} - (1 + e^{\mu X_{iFB}\beta})^{-1}$, $P_i(YN) = (1 + e^{\mu X_{iFB}\beta})^{-1} - (1 + e^{\mu X_{iLB}\beta})^{-1}$, $P_i(NN) = (1 + e^{\mu X_{iLB}\beta})^{-1}$.

[3] For the CV models, see Hanemann (1984), Hanemann et al. (1991), and Hanemann and Kanninen (1999). See also Cameron and Quiggin (1994) for a bivariate probit model for double-bounded CV, which accounts for the correlation between different responses by same individuals. Our focus is on comparing CV and CR data, and we address the issue of correlation between different responses from the same individuals by using random coefficient logit models that have a flexible error structure.

2.2. Random Coefficient Models

Although they frequently are applied to SP data, fixed coefficient logit models have some undesirable properties and assumptions. First, because of the independence of irrelevant alternatives property (McFadden 1974), they overestimate the joint probability of choosing close substitutes. Second, assuming i.i.d. stochastic terms ε_{ij} is restrictive, since in practice individual-specific factors are likely to influence the evaluation of all available alternatives and make the random terms correlated instead of independent. Random coefficient logit models have been proposed to overcome problems of the fixed coefficient logit models (Revelt and Train 1998; Train 1998; Layton 2000; see also Train 2000 for a comprehensive treatment of discrete choice modeling). A random coefficient model is specified similarly to the fixed coefficient model, except that the coefficients β are assumed to vary in the population rather than be fixed at the same value for each person. Utility is expressed as the sum of population mean *b*, an individual deviation η, which accounts for differences in individual taste from the population mean, and an unobserved i.i.d. stochastic term ε_{ij}. The total utility for person *i* from choosing alternative *j* is determined as $U_{ij} = X_{ij}b + X_{ij}\eta_i + \varepsilon_{ij}$, where $X_{ij}b$ and $X_{ij}\eta_i + \varepsilon_{ij}$ are the observed and unobserved parts of utility, respectively. Utility can also be written as $U_{ij} = X_{ij}(b+\eta_i) + \varepsilon_{ij}$, which shows how the previously fixed β now varies across people as $\beta_i = b + \eta_i$.

Although random coefficient models account for heterogeneous preferences via parameter η_i, individual taste deviations are neither observed nor estimated. The estimated model aims at finding the parameters that describe the distribution from which each β_i is drawn. Coefficients β vary across the population with density $f(\beta|\Omega)$, with Ω denoting the parameters of density. Since actual tastes are not observed, the probability of observing a certain choice is determined as an integral of the appropriate probability formula over all the possible values of β weighted by its density. The probability of choosing alternative *j* out of *m* alternatives can now be written as:

$$P_{ij} = \int \left[\frac{\mathrm{e}^{\mu X_{ij}\beta_i}}{\sum_{k=1}^{k=m} \mathrm{e}^{\mu X_{ik}\beta_i}} \right] f(\beta \,|\, \Omega) d\beta \,. \tag{4}$$

Equation (4) is the random coefficient extension of the conditional logit model (2). Extensions to the other models applied in this paper are straightforward and are not provided here. The multidimensional integral (4) typically must be simulated for estimation purposes. Train has developed a simulator that is smooth, strictly positive, and unbiased (Brownstone and Train 1999) and that can be easily modified for strictly negative or positive random coefficients. His method draws a random β_i, calculates the probability, and repeats the procedure many times. Using R draws of β_i from $f(\beta|\Omega)$ simulated probability is:

$$SP_{ij} = \frac{1}{R}\left[\sum_{r=1}^{r=R} \frac{\mathrm{e}^{\mu X_{ij}\beta_{ir}}}{\sum_{k=1}^{k=m} \mathrm{e}^{\mu X_{ik}\beta_{ir}}}\right]. \tag{5}$$

The CV and CR models typically are estimated using the method of maximum likelihood (the method of simulated maximum likelihood in the case of random coefficient models).

2.3. The Data Pooling Method

Data pooling involves combining separate sources of data, such as CV and CR, and estimating econometric models using the pooled data set. This enables comparing different data sources already at the estimation stage, which brings at least two benefits. First, likelihood ratio-based tests for data source invariance become available. Second, if data from different SP sources can be combined, the pooled econometric models provide a method to utilize all the information in the data collected.

The scale factor μ of the standard conditional logit model (2) is inversely related to the variance of the random component in the RUM. Using a single source of data, μ typically is set equal to one since otherwise β and μ could not be identified. The normalization of μ makes absolute values of the parameter estimates incomparable between different data sets; only the ratios of parameters are comparable across different sources of data (Swait and Louviere 1993).

Considering n sources of stated preference data, each data set $q = 1, ..., n$ provides a vector β_q of parameter estimates. Denoting the scale parameters of different data sources by μ_q, n vectors $\mu_q\beta_q$ of parameter estimates result. Pooling all n sources of data, $n-1$ scale parameters for different data sources can be identified. Fixing one scale factor, say $\mu_1 = 1$, the rest of the $n-1$ estimated scale parameters are inverse variance ratios relative to the reference data source (Hensher et al. 1999).

Using the data pooling method, the convergent validity of CV and CR data is easily tested. Denoting the vector of CV and CR parameter estimates by $\theta_{CV} = \mu_{CV}\beta_{CV}$ and $\theta_{CR} = \mu_{CR}\beta_{CR}$, the total unpooled (unrestricted) likelihood function for the combined data is:

$$LL(unpool) = LL_{CV}(Y_{CV} \mid X_{CV}, \theta_{CV}) + LL_{CR}(Y_{CR} \mid X_{CR}, \theta_{CR}), \tag{6}$$

where LL_{CV} and LL_{CR} are the total likelihood functions for CV and CR data. Similarly, Y_{CV} and Y_{CR} stand for the CV and CR response matrices and X_{CV} and X_{CR} for the CV and CR scenario attributes. Restricting $\theta_{CV} = \theta_{CR} = \theta_{Pool}$, the pooled (restricted) likelihood function is expressed as:

$$LL(pool) = LL_{CV}(Y_{CV} \mid X_{CV}, \theta_{Pool}) + LL_{CR}(Y_{CR} \mid X_{CR}, \theta_{Pool})\,. \tag{7}$$

Standard likelihood ratio tests can be applied to accept or reject the imposed restriction. If the LR- test statistic –2[*LL(pool) - LL(unpooled)*] is greater than the critical χ^2 value, the hypothesis $\theta_{CV} = \theta_{CR}$ is not supported and the pooling hypothesis is rejected. However, if the hypothesis is not rejected under equal scale parameters, the two data sets can be considered similar and the absolute parameter estimates are comparable across the CV and CR data. This test is a strict test of pooling, since it tests for the invariance of both the coefficients and the random components of RUM between the CV and CR data. A less stringent test of pooling can be conducted, for instance, by fixing $\mu_{CV} = 1$, estimating μ_{CR}, and testing for the equality of parameters to a multiplicative constant. If this restriction is not rejected, the data can be considered generated by the same taste parameters but have scale differences.

3. The Survey

Data were collected using a mail survey sent out in spring 1999 to 1,740 Finns between 18 and 75 years of age. The sample was drawn randomly from the official census register of Finland and is therefore a representative sample of the target population. The sample was divided randomly into two sub-samples of 840 and 900 respondents. The first sub-sample received a double-bounded CV questionnaire and the second sub-sample received a CR questionnaire.

WTP was measured for three alternative biodiversity hotspot conservation programs: increasing conservation from the currently protected 120,000 hectares to: (1) 275,000 hectares; (2) 550,000 hectares; and (3) 825,000 hectares. The current policy (status quo) protects 10 percent of all biodiversity hotspots; the new alternatives correspond to 25 percent, 50 percent, and 75 percent protection of hotspots (protecting between 1.4 and 4.2 percent of all forests in Finland). Table 1 explains, similarly to our questionnaires, the different conservation alternatives and their extent in relation to the total areas of forestland and biodiversity hotspots in Finland.

Table 1. Protection Alternatives Presented to the Respondents

Alternative	Hectares protected	Percentage of all forests	Percentage of all hotspots
Status quo	120,000 ha	0.6%	10%
Program 1	275,000 ha	1.4%	25%
Program 2	550,000 ha	2.8%	50%
Program 3	825,000 ha	4.2%	75%

In designing the survey, special attention was paid to making the conservation policy scenarios relevant and credible. An easy-to-read, one-page section in the questionnaire explained different conservation programs and their details. Proposed new programs were described as extensions to an already existing conservation program that uses incentive payments to encourage non-industrial private forest owners to set aside biodiversity hotspots. Although this program already had been discussed in the national news media and respondents were therefore more familiar with it than with a completely new policy program, the questionnaire described the current program in detail. While designing the survey, questionnaire versions went through several rounds of modifications and reviews by experienced SP practitioners as well as other economists, foresters, and ecologists with expertise in survey methods or biodiversity conservation. After the expert comments were incorporated, the questionnaires were tested by personal interviews and a pilot survey (n = 100) and modified on the basis of the results.

The questionnaires started with questions about the respondents' attitudes on different aspects of forest and public policies. The next section of the questionnaire described the forest management and current conservation situation in the country, followed by the valuation questions. The questionnaire concluded with questions on the respondents' socioeconomic background.[4]

Figure 1 illustrates the split-sample design employed in the survey. The CV sample was divided randomly into two equal sub-samples. The sub-sample A was asked for WTP for Programs 1 and 2 (275,000 and 550,000 hectares), whereas the sub-sample B was asked for WTP for Programs 2 and 3 (550,000 and 825,000 hectares).[5] So every CV respondent was asked two

---

[4] Complete questionnaires are available from the corresponding author upon request.

[5] As a reviewer pointed out, using a mail-in survey may change the nature of the double-bounded CV question by removing the surprise component, which is apparent, for example, when using one-on-one interviews. However, using a mail-in format does not inevitably limit the generality of our results. For example, Banzhaf et al. (2004) find no differences between double-bounded responses from a mail-in survey and an interactive internet-based survey,

consecutive WTP questions, and WTP for Program 2 was asked systematically in either the first or second WTP question, depending on the respondent's sub-sample. Therefore, our survey includes a split-sample test for ordering effects, which we return to later in the paper.

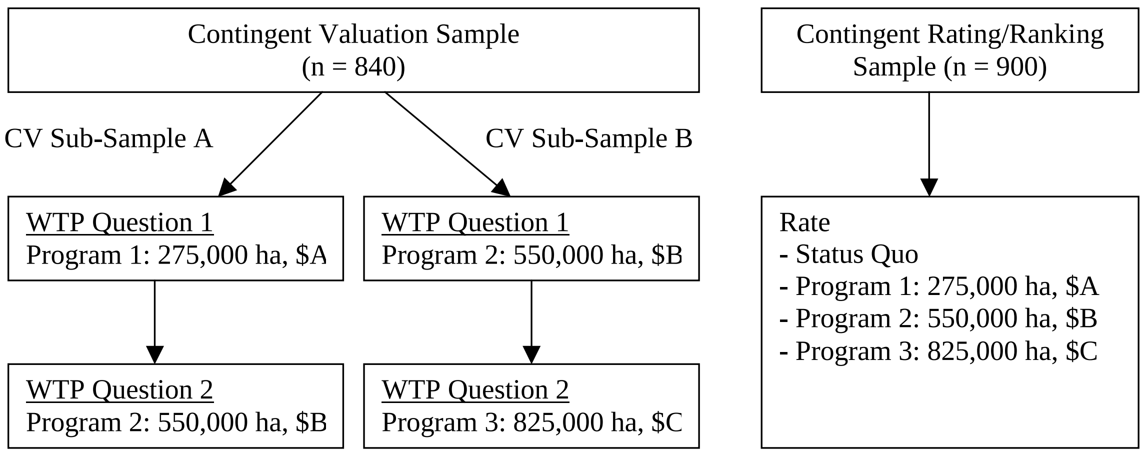

Figure 1. Split-Sample Survey Design

The CR questionnaire described the status quo and the three new programs (275,000, 550,000, and 825,000 hectares) to each respondent exactly in the same way as the CV questionnaires. Respondents were then asked to rate the status quo and three new programs on a scale from 0 to 10 so that the highest rating is given to the alternative they prefer (and would like to get implemented). In the CR survey, status quo and three new programs were evaluated simultaneously, not by asking three consecutive pair-wise comparisons involving status quo and a new program. An important decision about how to elicit preferences in the CR survey was made during the survey design. In one-on-one interviews conducted at different stages of questionnaire development, we experimented with asking respondents to either rate or rank the alternative programs. The respondents considered the rating of alternative programs an easier

---

which includes the surprise component. Also, follow-up questions have been applied earlier by using a variety of survey modes: telephone surveys, one-on-one interviews, and mail-in surveys. For example, Cameron et al. (2002) use a multiple-bounded mail-in CV survey. Herriges and Shogren (1996) use a mail-in double-bounded CV. Similar to their findings, we received mostly consistent and full response patterns to the double-bounded question. When some respondents responded to both follow-up questions, they did so in a consistent manner.

task than ranking them,[6] and we therefore used the rating approach in the final survey. To ensure that individual preference rankings of the proposed alternatives can be constructed based on the responses, we instructed respondents to state ratings, which reflect the preference order of the proposed programs.[7]

Besides the overall similarity of the CV and CR surveys, we avoided producing differences between the results from the two methods by using a similar bid vector design in both surveys. The CV and CR surveys used similar variations of the bids and the same bid amounts to the degree that is was possible given the overall questionnaire design. We also used an overlapping bid design that assigned the same bid amounts systematically to different respondents and different levels of conservation. The bids were described as total annual costs that the household of the respondent would have to pay for the proposed program. The bid vector in the CV survey consisted of first bids between \$4 and \$450, low follow-up bids between \$2 and \$230, and high follow-up bids between \$8 and \$690.[8] In the CR versions, the current policy (status quo) was described as a policy alternative that would not result in any additional costs to the respondent's household. The three new programs were assigned bids between \$4 and \$650. Within each CV and CR questionnaire version, policy programs with higher conservation percentages were assigned higher bids than programs with lower conservation percentages, thereby making the policy alternatives and their suggested costs credible in respondents' view.[9]

The final survey consisted of 29 different questionnaire versions, of which 14 used the CV and 15 used the CR. Except for the valuation question, the CV and CR questionnaires were exactly the same. The survey was mailed out in May 1999. A week after the first mailing, the whole sample was sent a reminder card. Two more rounds of reminders with a complete questionnaire were sent to non-respondents in June and July. The CV and CR surveys resulted in

---

[6] Some respondents felt that stating ratings gave them a better capacity to evaluate the proposed policy programs than rankings did, allowing them to express themselves as particularly strongly in favor of (or opposed to) certain programs(s). Obviously, for the purposes of our analysis, only the ordinal ranking of programs is used.

[7] Specifically, the questionnaire asked the highest rating to be given to the program that the respondent would like to have implemented. In addition, we asked ratings to express how valuable the respondent considers each program relative to its cost.

[8] Bids were assigned in local currency at the time of the survey, Finnish markka (FIM = US\$0.15).

[9] Allowing for positive correlation between bid levels and conservation levels was necessary based on feedback from the pilot study and interviews conducted while testing the early questionnaire versions. To avoid potential identification problems due to correlated design, we systematically varied the bid increases for higher conservation levels across different questionnaire versions.

very similar response rates of 48.9 percent and 50 percent, respectively.[10] After cleaning the data for missing responses to valuation questions, there were 308 observations for the CV data and 270 for the CR data.

4. Estimation

We first investigated whether CV responses were affected by question order. As explained, each CV respondent was asked two consecutive CV questions so that half of the CV sample evaluated Programs 1 and 2, while the other half evaluated Programs 2 and 3 (programs evaluated systematically in this order). We estimated a double-bounded CV model with program-specific dummies specified as: $V_{ij}= \beta_1 BID_i + \beta_{P1}P1 + \beta_{P2a}P2A + \beta_{P2b}P2B + \beta_3 P3$. Dummies $P1$ and $P3$ indicate Programs 1 and 3, and dummies $P2A$ and $P2B$ indicate Program 2 in the first and second CV question. BID is the cost of program j to person i. A LR-test between the unconstrained CV model, which estimates β_{P2a} and β_{P2b}, and the constrained model, which constraints $\beta_{P2a} = \beta_{P2b}$, results in a LR-test statistic value 0.03. This suggests that respondents' evaluations of Program 2 are independent of the question order, providing split-sample evidence for the absence of ordering effects in these data.

The CR data with the ratings of alternatives were first transformed into rankings. The derived rankings utilize only the ordinal preferences. Respondents with, for instance, ratings sequences (4, 3, 2, 1) and (10, 9, 3, 1) for the four policy alternatives provide responses with the same preference ordering A > B > C > D.[11]

Several alternative specifications for the indirect utility function were estimated.[12] Each specification uses status quo as the reference level of utility, which is normalized to zero. Alternative models specified the valuation function for conservation either as a continuous linear, logarithmic, quadratic, or piecewise linear using dummies for each conservation level as in Layton and Brown (2000). A non-nested model selection criterion, the likelihood dominance criterion (LDC) of Pollak and Wales (1991), was then employed in selecting a preferred specification. The LDC selected the quadratic model as the preferred specification for these data. Non-nested models selection results

---

10 Responses frequencies of the CV and CR surveys are statistically indifferent. A χ^2-test statistic for the difference of the CR response rate from the overall mean response rate equals 0.056, which is statistically highly insignificant (with one degree of freedom).

11 In building the rankings data, observations with ties or missing ratings were removed. On average, ties were observed in eight percent of comparisons. When combined with simulation estimation, modeling ties using the frameworks presented by either Boyle et. al. (2001) or Layton and Lee can be computationally demanding and is left as an objective for future work.

12 All the models were programmed and run in GAUSS.

are not reported in detail here, but their results are available in Siikamäki (2001). The observed part of the RUM is then estimated as:

$$V_{ij} = \beta_{BID} BID_{ij} + \beta_{Percent} PERCENT_{ij} + \beta_{SqrPercent} SQRPERCENT_{ij}, \tag{8}$$

where *BID* is the cost of program j to person i, *PERCENT* is the percentage increase in the biodiversity hotpots conserved under alternative j, and *SQRPERCENT* is its square.[13]

Previous applications typically have modeled either the *BID* or other alternative characteristics as random coefficients, not both. With these data, heterogeneity of preferences for conservation levels could be observed and *PERCENT* and *SQRPERCENT* were therefore modeled as normally distributed random coefficients.[14] On the other hand, the heterogeneity in responses is often related to the *BID* coefficient. Since it represents the negative of the marginal utility of income, it was estimated as a strictly negative Rayleigh-distributed random coefficient, $BID_{RAYLEIGH}$. Justification for using a Rayleigh distribution is discussed more in the following.

To allow for an estimable, bounded random parameter on *BID*, we propose the use of a single-parameter distribution. This has the advantage of being potentially more flexible than a degenerate distribution (e.g., fixed coefficient) but far easier to estimate in practice than a lognormal coefficient, especially with other random parameters in the model. To be useful as a distribution for the price coefficient, the distribution needs to be strictly negative (or strictly positive after entering price as its negative). More importantly, the distribution needs to be bounded away from zero or else its inverse will have no finite mean, implying that WTP will have no finite mean. An exponential distribution for instance, while being defined by a single parameter, has an inverse with no finite mean. A second criterion is the ease of drawing the random variable from a set of underlying random numbers that can be held fixed throughout the iteration of the maximum likelihood optimization program. A good candidate distribution is the Rayleigh distribution. As shown in the Appendix, which summarizes the properties of the Rayleigh distribution, a Rayleigh distributed *BID* is easy to draw by inversion, and the mean of its reciprocal is finite. How well a single-parameter distribution facilitates the estimation of

[13] While the quadratic form indicates declining WTP at some point, we use it as it fits best according to the LDC. That is, we prefer not to use a sub-optimal form for testing the pooling of the two types of data as it might lead to erroneous conclusions.

[14] No correlation parameters between random coefficients were estimated. Models with correlation between random parameters could not be identified in practice.

random parameter models must be determined empirically. As our editor pointed out, as a purely mathematical matter, a Rayleigh-distributed random coefficient does not necessarily lead to an improvement in the maximized log-likelihood value of the estimated model. This is because a random coefficient model using a single-parameter distribution estimates the same number of parameters as a fixed coefficient model.[15]

5. Results

5.1. Separate Contingent Valuation and Contingent Ranking Models

Table 2 reports both fixed and random coefficient logit (FCL and RCL) model results for separate CV and CR data. We show the results of a variety of models, including both single- and double-bounded CV models, to later highlight the systematic tendencies in the convergent validity of CV and CR data. Single-bounded CV models are estimated using only the responses to the first bid.[16] The CR models explain the ranking of four alternatives that respondents were presented with. "One Rank CR" is the standard conditional logit model for the first rank (the preferred alternative), "Two Ranks CR" is a rank-ordered logit model for the first and second ranks, and "Three Ranks CR" is a rank-ordered logit model for the full ranking of the presented policy alternatives.

[15] A referee helpfully points out that when flexibility is a driving force in choosing a distribution, there are other useful approaches for introducing bounded random coefficients. For example, Train (1998) estimates strictly negative coefficients as log-normally distributed. A disadvantage of the lognormal random coefficients is that they are often very hard to estimate and identify (e.g., Train 2001, 2003). Layton (2000) used the lognormal, but it was the only random parameter. In our application, which requires estimation of a variety of pooled models, the estimation of log-normally distributed *BID* coefficient is not possible for all the necessary models. Train and Sonnier (2005) develop and discuss the use of Johnson's (1949) S_B distribution in a Markov Chain Monte Carlo routine. The S_B distribution is easy to draw from, can be generalized to have up to four parameters, and has a finite mean for the reciprocal distribution. However, the reciprocal distribution mean does not have a known closed form. While the approach is very promising, in practice the extra parameters are likely difficult to identify using simulated maximum likelihood. One can think of other four-parameter distributions that could be used but again identification in practice is important. For this reason, we find that using a single-parameter distribution has merit.

[16] We present single-bounded CV results mostly for comparison. A reviewer suggested using a bivariate probit model (Cameron and Quiggin 1994) to contrast the WTP distributions from first and second responses. Bivariate probit models, which allow for unequal variances of WTP distributions but constrain their means to be equal, do not reject the equality of the mean WTP from the first and second responses. The LR-test statistics (p-values in brackets) for Program 1, 2a, 2b, and 3 are 6.42 (0.04), 0.18 (0.91), 1.05 (0.59), 0.36 (0.83), respectively. These results also indicate that the correlation between the WTP distributions is significantly less than unity.

Table 2. Estimates from Separate CV and CR Models*

Model	SB CV		DB CV		One Rank CR		Two Ranks CR		Three Ranks CR	
Coefficient	FCL	RCL	FCL	RCL	FCL	RCL	FCL	RCL	FCL	RCL
BID	-0.1959 (8.982)		-0.3099 (19.730)		-0.0860 (4.746)		-0.0752 (5.803)		-0.0557 (5.232)	
$BID_{RAYLEIGH}$		-0.6354 (3.687)		-1.1390 (9.381)		-0.0752 (3.311)		-0.1905 (2.901)		-0.1096 (2.311)
Means										
PERCENT	0.7449 (7.717)	3.2989 (3.575)	0.7895 (10.306)	2.1493 (10.039)	0.1831 (1.677)	0.2268 (1.056)	0.4180 (5.290)	2.3525 (5.239)	0.6925 (10.124)	.7012 (7.458)
SQRPERCENT	-0.1116 (6.668)	-0.5669 (3.381)	-0.1055 (7.787)	-0.2515 (8.461)	-0.0257 (1.671)	-0.0350 (0.851)	-0.0667 (5.875)	-0.4737 (4.882)	-0.1075 (11.110)	-0.3217 (9.832)
Deviations										
PERCENT		1.5776 (3.235)		0.7081 (6.871)		-0.0350 (0.135)		2.1215 (5.131)		.8121 (8.383)
SQRPERCENT		0.0056 (0.061)		0.0194 (0.802)		0.0471 (0.272)		0.1765 (2.632)		0.0079 (0.285)
LL	-338.78	-299.79	-926.55	-772.07	-347.82	-348.73	-608.78	-479.99	-746.98	-545.31

* Absolute t-statistics in brackets

The fixed coefficient models result in statistically significant estimates for all the parameters. Further, the *BID* parameter gets an expected negative sign in all the fixed coefficient models.[17] In addition, *PERCENT* and *SQRPERCENT* coefficients get positive and negative estimates, suggesting that the value of conservation is first increasing and then decreasing in the conservation area. In comparison to the fixed coefficient models, estimating random coefficients improves the statistical performance of models for CV data and the CR data for two and three ranks. Interestingly, the CR model for the first rank only with random coefficients performs statistically slightly worse than the fixed parameter model.[18] This suggests that a model with Rayleigh-distributed coefficients is potentially, but not necessarily, more flexible than a model with fixed coefficients. To investigate whether Rayleigh-distribution has the potential for estimating random coefficients, we estimated various models with Rayleigh-distributed *BID* as the only random

[17] *BID* and *PERCENT* are scaled for estimation by dividing the variable values by 100 and 10. Also, since $BID_{RAYLEIGH}$ parameter is positive by construction, random coefficient models enter bid values as their negative.
[18] We checked the result by numerous runs that used different starting values and iterative algorithms. Apparently for this model, a random parameter formulation with a Rayleigh-distributed BID does not provide improvements in the log-likelihood.

coefficient. For example, a CR model for two ranks with a Rayleigh-distributed *BID* as the only random coefficient results in a maximized log-likelihood value of -599.6. Comparing this to the log-likelihood value of the parallel, fixed coefficient model, -608.8, and taking into account that both models estimate the same number of parameters, shows that estimating a Rayleigh-distributed random coefficient can significantly improve the maximized log-likelihood value relative to the fixed coefficient models. Other CR and CV models give similar results when estimated with a Rayleigh-distributed bid as the only random coefficient.

The estimated Rayleigh-distributed coefficients are perhaps best compared to the fixed coefficients by using the formula for the mean of the Rayleigh-distributed *BID* parameter from the Appendix. The estimated $BID_{RAYLEIGH}$ parameter of the SB and DB CV models translates to means of -0.796 and -1.427 for *BID* and $BID_{RAYLEIGH}$ estimates for CR models for one, two and three ranks into *BID* means of -0.094, -0.239, and -0.137, respectively.

The random coefficient models for CV data and for CR data on two and three ranks generate a statistically significant estimate of *PERCENT* variable, which suggests that preferences regarding the conservation level are heterogeneous. The curvature parameter, *SQRPERCENT*, has a statistically significant deviation estimate only in the CR model for two ranks, but the estimate for the mean of the *SQRPERCENT* parameter is significant for all models except the “One Rank CR” model. Despite its insignificant *SQRPERCENT* parameter, the LDC suggests that the quadratic “One Rank CR” model with random coefficients is still preferred over the models that use either a linear or logarithmic specification of the utility from conservation.

5.2. Pooled Models

Table 3 reports the results of the pooled CV and CR models by categorizing models by the number of ranks estimated in the CR part of the pooled model. Although earlier analysis already suggested that the random coefficient formulation is favored for modeling these data, results of both fixed and random coefficient models are presented to demonstrate the effect of alternative modeling choices on the pooling tests. We focus on testing the “weak” pooling hypothesis H_0: $\beta_{CV} = \mu_{CR}\beta_{CR}$, according to which the vector of estimated parameters is common up to a multiplicative constant between the CV and CR data. Each pooled model therefore estimates a relative scale factor μ_{CR}, which accounts for differences in the variance of the stochastic term of the RUM between the CV and CR data. If the variance of the stochastic term is similar in both the CV and CR data, the estimate of μ_{CR} is not statistically different from one. Since the scale factor is inversely related to the variance of the stochastic component of the RU-model, an estimate $\mu_{CR} < 1$ suggests that the CR data is noisier than the CV data and $\mu_{CR} > 1$ the opposite.

Table 3. Estimates from Pooled Models*

Pooled Models	One Rank CR				Two Ranks CR				Three Ranks CR			
	SB CV		DB CV		SB CV		DB CV		SB CV		DB CV	
Coefficient	FCL	RCL	FCL	RCL	FCL	RCL	FCL	RCL	FCL	RCL	FCL	RCL
BID	-0.2001 (9.206)		-0.3100 (19.816)		-0.1879 (8.930)		-0.3067 (19.680)		-0.1560 (7.291)		-0.2989 (19.196)	
$BID_{Rayleigh}$		-0.6575 (3.708)		-1.1350 (9.375)		-0.5014 (2.846)		-0.9171 (9.264)		-0.4081 (2.889)		-0.8897 (9.679)
Means												
PERCENT	0.7103 (7.427)	3.1104 (3.661)	0.7857 (10.425)	2.1547 (10.090)	0.7871 (8.705)	3.7493 (2.897)	0.8597 (12.029)	2.4409 (10.339)	0.8758 (10.045)	3.1389 (2.899)	0.9601 (13.574)	2.4011 (10.717)
SQRPERCENT	-0.1048 (6.424)	-0.5192 (3.482)	-0.1051 (7.932)	-0.2540 (8.570)	-0.1210 (7.997)	-0.6941 (2.819)	-0.1198 (9.621)	-0.2974 (9.446)	-0.1380 (9.484)	-0.5407 (2.849)	-0.1383 (11.291)	-0.2846 (10.204)
Deviations												
PERCENT		1.4653 (3.339)		0.7060 (6.873)		2.4451 (2.914)		1.2133 (8.874)		2.4874 (2.951)		1.2315 (9.118)
SQRPERCENT				0.0195 (0.809)		0.2683 (0.1121)		0.0494 (1.809)		-0.0342 (0.645)		0.0059 (0.284)
μ_{CR}	0.4109 (4.928)	0.0987 (2.858)	0.2916 (5.659)	0.0686 (4.289)	0.4506 (6.517)	0.6545 (2.652)	0.2952 (7.923)	0.8117 (5.878)	0.4956 (6.831)	0.5739 (2.811)	0.2758 (8.950)	0.8479 (7.492)
LL pooled	-687.66	-649.99	-1,274.47	-1,121.52	-948.20	-794.06	-1,540.71	-1,335.42	-1,103.02	-866.00	-1,708.82	-1,406.27
LL unpooled	-686.60	-648.45	-1,274.37	-1,120.73	-947.56	-779.78	-1,535.33	-1,252.06	-1,085.76	-845.10	-1,673.53	-1,317.38
LR-test	2.12	3.10	0.20	1.58	1.28	28.56	10.76	166.73	34.52	41.80	70.58	177.78
P-value	0.34645	0.54123	0.90483	0.81238	0.52729	0.00001	0.00461	0.00000	0.00000	0.00000	0.00000	0.00000
Convergent Validity	Accepted	Accepted	Accepted	Accepted	Accepted	Rejected	Rejected	Rejected	Rejected	Rejected	Rejected	Rejected

* Absolute t-statistics in brackets

LR tests are employed in accepting or rejecting the pooling hypothesis; the respective LR-test statistics are reported in the last row of the Table 3. The test statistic is distributed χ^2, with degrees of freedom equal to the difference in the number of estimated parameters between pooled and unpooled models.[19] To facilitate comparisons across different models with different degrees of freedoms, we present also the p-values of each LR-test statistic in Table 3.

Each model combining CV and “One Rank CR” estimates a statistically significant scale parameter μ_{CR}, which yields a value less than one. All LR-test statistics (2.12, 3.10, 0.20, and 1.58) are smaller than their respective critical values. Both fixed and random coefficient models provide support for accepting the pooling of CV and “One Rank CR.”

“Two Ranks CR” models combined with CV models reject the pooling hypothesis, except when estimating a fixed coefficient model with a single-bounded CV-part of the pooled model (LR-test statistics 1.28). The same model with random coefficients results in a LR-test statistics 28.56, suggesting a clear rejection of the pooling hypothesis. Fixed and random coefficient models, which include “Two Rank CR” models and double-bounded CV models, result in LR-test statistics 10.76 and 166.73, and reject the pooling hypothesis.

Combining the “Three Ranks CR” models with CV models also rejects the pooling hypothesis, with yet stronger statistical support than the pooled models including a “Two Ranks CR” model.

The estimated relative scale parameters for the CR data are statistically significantly less than one in all estimated models. This suggests that the variance of RUM random term is higher in the CR data than in the CV data. In addition, keeping everything else constant, the estimates of relative scale parameters increase when more ranks are included in the CR model. This tendency is logical since additional ranks increase the amount of information on individual preferences utilized in estimation, hence, decreasing the variance of the RUM error term.

Several past studies suggest (for example, Hausman and Ruud 1987; Ben-Akiva et al. 1991; Layton 2000; Boyle et al. 2001) that preferences are not consistent across different ranks. We investigated the consistency of ranks using the approach in Hausman and Ruud (1987) and, in summary, found that fixed coefficient models reject the consistency of ranks for both two and three ranks. We do not extend the Hausman and Ruud (1987) approach to random coefficient

---

[19] Degrees of freedom for pooled fixed and random coefficient models equal to 2 and 4, respectively. The critical values with 2 degrees of freedom at five percent and one percent significance levels are 5.99 and 9.21; the respective critical values for 4 degrees of freedom are 9.49 and 13.28.

models, but the fixed coefficient model results and the pooling results lend some support to the suggestion that the inconsistency of rankings may cause the lack of convergent validity between the CV data and CR data for more than the preferred choice.

The pooled model results resemble closely the results of separate CV models. This is primarily a consequence of normalizing the CV scale parameter to one in the pooled models. Accordingly, the pooled models estimate coefficients to the scale of the CV coefficients, which is then reflected in the similarity of the results of pooled and CV models.

Although not reported in Table 3, a "strong" pooling hypothesis H_0: $\beta_{CV} = \beta_{CR}$ also was examined. It was tested by estimating models with a restriction $\mu_{CR} = 1$, that is, by imposing equal variance of the stochastic term of the RUM between the CV and CV models. Using this completely pooled model, the pooling of the CV and CR data is uniformly and strongly rejected. Complete invariance of the CV and CR data is therefore uniformly rejected. In random coefficient models, yet another pooling possibility is to allow CV and CR parameter distributions to have the same means but different variances. We tested for the pooling of the CV and Two- and Three Ranks-CR models under this "weakest" criterion for pooling. These tests continue to reject the pooling of CV and Two-and Three Ranks-CR models.[20]

6. Willingness to Pay

Since our results consist of a variety of different models, the question becomes which results are preferred and chosen for further purposes, such as policy evaluation. We use the acceptance of pooling of CV and CR data as one criterion for selecting models for policy evaluation. The models that successfully pool the CV and CR data are less dependent on the chosen survey method and therefore can be considered more general than separate CV or CR models. All fixed and random coefficient models involving "One Rank CR" are accepted by the pooling criterion. We then continue by setting aside fixed coefficient models as more restrictive and poorer fitting than their random coefficient counterparts. We then select the pooled model comprising a "One Rank CR" model and a double-bounded CV model as the preferred model for

[20] Also, we earlier have (Siikamäki 2001) tested for the pooling of double-bounded CV and CR data using a piecewise linear utility specification, which is perhaps the most prominent alternative specification to the quadratic specification used here. Estimating a full suite of fixed and random coefficient models with a piecewise linear specification (involving program specific constants), we confirmed that these models result qualitatively in the exact same conclusions as the quadratic models. That is, the piecewise models reject and accept the pooling of CV and CR data in the same way the quadratic models do. Nevertheless, the LDC criterion prefers the quadratic specification.

further evaluation, since this model both utilizes the data comprehensively and satisfies the pooling criterion.

The WTP for policy alternative x_j is calculated as $\partial V / \partial x_j * (\partial V / \partial y)^{-1}$, where the $\partial V / \partial x_j$ is the utility change from reaching some conservation level, and $(\partial V / \partial y)^{-1}$ is the inverse of the marginal utility of income, that is, the inverted *BID* parameter. The estimates for *PERCENT* and *SQRPERCENT* parameters are used in calculating a point estimate for the utility change from increasing conservation from status quo to each program alternative. A money measure of the change in utility is then obtained by multiplying it by the inverted marginal utility of income.

Calculating the mean WTP estimates from the fixed coefficient model results is straightforward: The fixed coefficient estimates equal their expectation, and with linear-in-income specification, the ratio of estimates for the utility changes $\partial V / \partial x_j$ and $\partial V / \partial y$ expresses the mean WTP. For random coefficient models, obtaining WTP estimates requires calculating the expectation of the inverted Rayleigh distributed *BID*. The expectation of the inverse of a Rayleigh random variate is provided in the Appendix. The mean estimates of WTP are reported in Table 4, as well as their confidence intervals, which were simulated by using the method of Krinsky and Robb (1986). The closed-form expressions for the mean of the reciprocal Rayleigh distribution facilitate the simulation of the confidence intervals. For the sake of comparison, we also report WTP estimates for the corresponding fixed coefficient model.

Table 4. Willingness-to-Pay Estimates [a]

	Program 1			Program 2			Program 3		
Model	CI 5[b]	Mean	CI 95[b]	CI 5	Mean	CI 95	CI 5	Mean	CI 95
RCL, preferred	*37*	45.4	*54*	*59*	77.6	*97*	*19*	55.5	*92*
FCL, alternative	*39*	46.8	*55*	*53*	72.5	*92*	*0*	32.8	*79*

[a] All estimates in $US (1999), which exchange at 1:6.5 for Finnish markka.

[b] Confidence intervals (the lower limit CI 05 and the upper limit CI 95) are reported in *italics*, rounded to integers using $0 as the lower limit. Confidence intervals are based on 10,000 draws from the estimated parameters and their variance-covariance matrix. No truncation of WTP was employed in the simulation.

The WTP estimates are similar between the fixed and random coefficient models. The WTP for Program 1 (25 percent conservation) in the fixed coefficient model is about $47, whereas the random coefficient model produces a WTP estimate of about $45. The estimates of WTP for Program 2 (50 percent conservation) are also alike ($73 and $78). The models behave somewhat differently in estimating the WTP for Program 3 (75 percent conservation). The fixed coefficient model gives an estimate of $33, whereas the random coefficient model produces a somewhat higher estimate, $56. Both models estimate highest mean WTP for Program 2, although the confidence intervals of the WTP for Program 3 overlap the mean WTP for Program 2.

7. Conclusion

This paper considers and compares the convergent validity of different SP response formats using traditional fixed coefficient and more recent random coefficient models. We find that CV and CR methods produce statistically indistinguishable estimates of WTP when similar changes of environmental quality are considered. Convergent validity of CV and CR methods is attained when the estimated CV and CR parameters are allowed to have scale differences and only the preferred CR alternative is modeled. Therefore, our results suggest that as long as modeling preferences does not go beyond what people usually do in the market—that is, choose their preferred alternative—the CV and CR methods yield converging results. However, when the CR data includes rankings of alternatives beyond choosing only the preferred alternative, the CR and CV model results start to diverge and become inconsistent. This suggests that the inconsistency of rankings may cause the lack of convergent validity between the CV data and CR data for more than the preferred choice.

In terms of econometrics, this paper considers and selects a nonlinear functional form for the indirect utility function. We also introduce the Rayleigh distribution for the price coefficient, which possesses a number of desirable properties useful for SP valuation studies. While the literature on random coefficient models has grown significantly during recent years (e.g., Hensher and Green 2003; Train and Sonnier 2005) and now offers a number of possible multi-parameter distributions, in many empirical studies it is difficult in practice to identify multi-parameter distributions. This is especially likely to cause difficulties in applications such as ours that require successful estimation of a large number of comparable models.

We find that random coefficient models do not appear to provide a methodological fix for the pooling of CV and CR data. In fact, the flexibility of the random coefficient models may help to identify differences between different sources of data that more rigid, fixed coefficient models do not discover. Summing up, differences in respondents' behavior in ranking tasks seem to be the primary contributor to the differences between the CV and CR methods, not the inflexibility of the econometric methods typically used for modeling the responses.

What are the broader implications of our findings? When both CV and CR methods recover the same preferences (as long as preferred choices and similar changes in environmental quality are considered), the choice between them does not affect the estimated welfare effects. For this reason, the researcher may focus on other important considerations when choosing between different SP methods. Each valuation problem is different and may call for a different method. Considerations, such as how well different methods facilitate communicating specific policy problems to the respondents, eliciting their preferences, or using benefits transfer techniques for estimating benefits from different policy configurations or geographical locations, may influence what method the researcher finds suitable for examining a particular valuation problem. Finally, if the researcher is especially interested in ensuring that the welfare effects are not driven by the choice of SP method, then the approach to the elicitation and estimation of preferences demonstrated in this paper, which uses multiple elicitation methods and estimates WTP that is common across them, may be helpful in conducting a study.

The analysis in this paper highlights only some of the prospective uses of pooled econometric models in examining SP survey methods. The driving force of differences between alternative SP methods could be examined further using the same framework, for instance, by modeling the response incentives in different SP surveys in the fashion of Alberini et al. (1997). In addition, a pooled econometric framework, in combination with varying experimental design treatments (following the approaches taken by DeShazo and Fermo 2002 and Hensher 2003), could be applied to investigate how experimental design attributes (such as number of alternatives, attributes, and so forth) influence the convergent validity of alternative SP methods. If the differences between alternative SP data sources can be controlled for in estimation, the pooled econometric models can be used in estimating WTP using several SP data sources.

Appendix

Properties of a Rayleigh Distributed Random Variable and Its Reciprocal

A random variable, *x*, has a Rayleigh distribution if its probability density function (pdf), *f*(*x*) is

$$f_x(x) = \left(x^2 / b^2 \right) \exp\left[-\left(x^2 / 2b^2 \right) \right]. \tag{A1}$$

Its distribution function is

$$F_x(x) = 1 - \exp\left[-\left(x^2 / 2b^2 \right) \right], \tag{A2}$$

where *b* is the only parameter and must be greater than 0. The range of *x* is $0 < x < \infty$. Figure A1 plots the Rayleigh pdf for different values of *b*. The mean, variance, median and mode of the Rayleigh-distributed *BID* are $b(\Pi/2)^{1/2}$, $(2-\Pi/2)b^2$, $b(\log 4)^{1/2}$, and *b*, respectively (Johnson *et al.* 1994). To draw a Rayleigh distributed *x* using a random uniform, *u*, invert the distribution function in (A2), and compute $x = (-2b^2 \ln(1-u))^{1/2}$.

Now let *y* be the reciprocal of *x*. Then the pdf of *y* can be found using standard results on the function of a random variable $y = h(x)$, which implies $x = g(y)$ where *h()* and *g()* are both invertible functions. The pdf of *y* is

$$f_y(y) = f_x(g(y)) \left| g'(y) \right|, \tag{A3}$$

where $|g'(y)|$ is the absolute value of the derivative of *g(y)*. In our case $g(y) = 1/y$, and $|g'(y)| = 1/y^2$. Applying (A3), we find that the density of *y* is

$$f_y(y) = \left(1 / (b^2 y^3) \right) \exp\left[-\left(1 / (2 b^2 y^2) \right) \right]. \tag{A4}$$

The range of *y*, like *x*, is $0 < y < \infty$. The distribution function of *y* is shown in (A5) below and is found by direct integration of (A4) over the range of *y*.

$$F_y(y) = \exp\left[-\left(1 / (2 b^2 y^2) \right) \right]. \tag{A5}$$

The median of *y* is found by setting (A5) equal to .5, and solving for *y*:

$$\text{Median}(y) = \sqrt{\frac{-1}{\left(2b^2 \ln(.5)\right)}}. \tag{A6}$$

To find E(*y*), use $E(y) = \int_0^\infty y f_y(y) dy$ and substitute $w = 1/\sqrt{2yb}$. After some algebra the following integral is obtained:

$$E(y) = \left(\frac{\sqrt{2}}{b}\right) \int_0^\infty \exp\left(-w^2\right) dw. \tag{A7}$$

The integral is $\sqrt{\pi}/2$ times the Gaussian error function. The Gaussian error function from 0 to infinity integrates to 1, so the mean of *y* is:

$$E(y) = \frac{\sqrt{\pi}}{b\sqrt{2}}. \tag{A8}$$

References

Adamowicz, W., J. Louviere, and M. Williams. 1994. Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. *J. Environ. Econom. Management* 26: 271–292.

Adamowicz, W., P. Boxall, M. Williams, and J. Louviere. 1998. Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation. *Amer. J. Agric. Econ. 80*: 64–75.

Alberini, A., B. Kanninen, and R. T. Carson. 1997. Modeling Response Incentives in Dichotomous Choice Contingent Valuation Data. *Land Econom.* 73: 309–24.

Azevedo, C. D., J. A. Herriges, and C. L. Kling. 2003. Combining Revealed and Stated Preferences: Consistency Tests and Their Interpretations. *Amer. J. Agric. Econ.* 85(3): 525–537.

Banzhaf, H. S., D. Burtraw, D. Evans, and A.J. Krupnick. 2004. Valuation of Natural Resources in the Adirondacks. Washington, DC: Resources for the Future.

Beggs, S., S. Cardell, and J. Hausman. 1981. Assessing the Potential Demand for Electric Cars. *J. Econometrics* 16: 1–19.

Ben-Akiva, M. E., T. Morikawa, and F. Shiroishi. 1991. Analysis of the Reliability of Preference Ranking Data. *Journal of Business Research* 23(3): 253–268.

Boxall, P., W. Adamowicz, J. Swait, M. Williams, and J. Louviere. 1999. A Comparison of Stated Preference Methods for Environmental Valuation. *Ecoloc. Econom.* 18: 243–253.

Boxall, P., J. Englin, and W. Adamowicz. 2003. Valuing Aboriginal Artifacts: A Combined Revealed-Stated Preference Application. *J. Environ. Econ. Management 45*: 213–230.

Boyle, K. J., T.H. Holmes, M.F. Teisl, and B.A. Roe. 2001. A Comparison of Conjoint Analysis Response Formats. *Amer. J. Agric. Econ.* 32: 441–454.

Brownstone, D., and K. Train. 1999. Forecasting New Product Penetration with Flexible Substitution Patterns. *J. Econometrics* Vol. 89: 109–129.

Cameron, T. A. 1992. Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods. *Land Econom.* 68: 302–317.

Cameron, T. A., and J. Quiggin. 1994. Estimation Using Contingent Valuation Data from a "Dichotomous Choice with Follow-Up" Questionnaire. *J. Environ. Econ. Management* 27: 218–234.

Cameron, T. A., G. L Poe, R. G. Ethier, and W. D. Schulze. 2002. Alternative Nonmarket Value-Elicitation Methods: Are Underlying Preferences the Same? *J. Environ. Econ. Management* 44(3): 391–425.

Chapman, R. G., and R. Staelin. 1982. Exploiting Rank Ordered Choice Set Data within the Stochastic Utility Model. *J. Marketing Res.* 19: 288–301.

Chen, H. Z., and S. R. Cosslett. 1998. Environmental Quality and Preference Estimation In Multinomial Probit Models: A Simulation Approach. *Amer. J. Agric. Econ.* 80: 512–520.

Cummings, R. G., S. Elliott, G. W. Harrison, and J. Murphy. 1997. Are Hypothetical Referenda Incentive Compatible? *J. Pol. Econ* 105:609–621.

Desvousges, W. H., and V. K. Smith. 1983. Option Price Estimates for Water Quality Improvements: A Contingent Valuation Study for Monongahela River. *J. Environ. Econ. Management* 14: 248–267.

Foster, V., and S. Mourato. 2002. Testing for Consistency in Contingent Ranking Experiments. *J. Environ. Econ. Management* 44: 309–328.

Haab, T. C., J-C. Huang, and J. C. Whitehead. 1997. Are Hypothetical Referenda Incentive Compatible? A Comment. *J. Pol. Econ*. 107:186–196.

Haab, T. C., and K. E. McConnell. 2002. *Valuing Environmental and Natural Resources: The Econometrics of Nonmarket Valuation*. Northampton, MA: Edward Elgar.

Halvorsen, B. 2000. Comparing Ranking and Contingent Valuation for Valuing Human Lives, Applying Nested and Non-Nested Logit Models. *Environ. Res. Econom.* 17: 1–19.

Hanemann, M. W. 1984. Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses. *Amer. J. Agric. Econ.* 66: 332–341.

Hanemann, M. W., J. Loomis, and B. Kanninen. 1991. Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation. *Amer. J. Agric. Econ.* 73: 1255–1263.

Hanemann, M. W., and B. Kanninen. 1999. The Statistical Analysis of Discrete-Response CV Data. In *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries,* edited by I. Bateman and K. Willis. Oxford: Oxford University Press, 302–442.

Hausman, J. A., and P.A. Ruud. 1987. Specifying and Testing Econometric Models for Rank-Ordered Data. *J. Econometrics* 34: 83–104.

Hensher, D. 2003. Revealing Differences in Willingness to Pay due to the Dimensionality of Stated Choice Designs: An Initial Assessment. Working paper, Institute of Transport Studies, University of Sydney.

Hensher, D., and W. H. Green. 2003. The mixed logit model: The state of practice. *Transportation* 30(2): 133–176.

Hensher, D., J. Louviere, and J. Swait. 1999. Combining Sources of Preference Data. *J. Econometrics* 89: 197–221.

Herriges, J. A., and J. F. Shogren. 1996. Starting Point Bias in Dichotomous Choice Valuation with Follow-Up Questioning. *J. Environ. Econom. Management* 30: 112–131.

Huang, J., T. Haab, and J. Whitehead. 1997. Willingness-to-Pay for Quality Improvements: Should Revealed and Stated Preferences be Combined? *J. Environ. Econom. Management* 34: 240–255.

Johnson, N. L. 1949. Systems of Frequency Curves Generated by Methods of Translation. *Biometrika* 36(1/2): 149–176.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. *Continuous Univariate Distributions*, 2nd edition. New York: John Wiley & Sons.

Kling, C. L. 1997. The Gains from Combining Travel Cost and Contingent Valuation Data to Value Nonmarket Goods. *Land Econom.* 73: 428–439.

Krinsky, I., and A. Robb. 1986. Approximating the Statistical Properties of Elasticities. *Rev. Econom. Stat.* 68: 715–719.

Layton, D. F. 2000. Random Coefficient Models for Stated Preference Surveys. *J. Environ. Econom. Management* 40: 21–36.

Layton, D. F., and G. Brown. 2000. Heterogeneous Preferences Regarding Global Climate Change. *Rev. Econom. Stat.* 82: 616–624.

Layton, D. F., and S. T. Lee. Forthcoming. From Ratings to Rankings: The Econometric Analysis of Stated Preference Ratings Data. In *Explorations in Environmental and Natural Resource Economics: Essays in Honor of Gardner M. Brown,* edited by R. Halvorsen and D. F. Layton. Northampton, MA: Edward Elgar Publishing.

Louviere, J. J., D. A. Hensher, and J. D. Swait. 2000. *Stated Choice Methods: Analysis and Applications*. Cambridge: Cambridge University Press.

Magat, W. A., W.K. Viscusi, and J. Huber. 1988. Paired Comparison and Contingent Valuation Approaches to Morbidity Risk Valuation. *J. Environ. Econ. Management* 15: 395–411.

McFadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academic Press, 105–142.

McFadden, D., and K. Train. 2000. Mixed MNL Models for Discrete Response. *J. Applied Econometrics* 15: 447–470.

Pollak, R. A., and T.R. Wales. 1991. The Likelihood Dominance Criterion: A New Approach to Model Selection. *J. Econometrics* 47: 227–242.

Ready, R., J. Whitehead, and G. Blomquist. 1995. Contingent Valuation When Respondents Are Ambivalent. *J. Environ. Econom. Management* 29: 181–197.

Revelt, D., and K. Train. 1998. Mixed Logit with Repeated Choice: Households' Choices of Appliance Efficiency Level. *Rev. Econom. Stat.* 80: 647–657.

Siikamäki, J. 2001. Discrete Choice Experiments for Valuing Biodiversity Conservation in Finland. Ph.D. dissertation, University of California, Davis.

Stevens, T. H., R. Belkner, D. Dennis, D. Kittredge, and C. Willis. 2000. Comparison of Contingent Valuation and Conjoint Analysis in Ecosystem Management. *Ecoloc. Econom.* 32: 63–74.

Swait, J., and J. Louviere. 1993. The Role of Scale Parameter in the Estimation and Use of Generalized Extreme Utility Models. *J. Marketing Res.* 10: 23–45.

Train, K. 1998. Recreation Demand Models with Taste Differences Over People. *Land Econom.* 74: 230–239.

Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Train, K. 2001. A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit. Department of Economics, University of California, Berkeley.

Train, K., and G. Sonnier. 2005 Mixed Logit with Bounded Distributions of Correlated Partworths. In *Applications of Simulation Methods in Environmental and Resource Economics*, edited by R. Scarpa and A. Alberini. The Netherlands: Springer, Dordrecht, 117–134.

Welsh, M. P., and G.L. Poe. 1998. Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach. *J. Environ. Econom. Management* 36: 170–185.