WORKING PAPER
1998-15

sustainable forest management network

réseau sur la gestion durable des forêts

How Accurate are Dichotomous Choice Contingent Valuation Welfare Measures when Agents have Heterogeneous Preferences?

A Network of Centres of Excellence
Un réseau de centres d'excellence

M. Rudd and G.C. van Kooten

For copies of this or other SFM publications contact:

Sustainable Forest Management Network
G208 Biological Sciences Building
University of Alberta
Edmonton, Alberta, T6G 2E9
Ph: (780) 492 6659
Fax: (780) 492 8160
http://www.biology.ualberta.ca/sfm

How Accurate are Dichotomous Choice Contingent Valuation Welfare Measures when Agents have Heterogeneous Preferences?

by

Murray Rudd

Workshop in Political Theory and Policy Analysis
Indiana University
513 North Park
Bloomington, Indiana 47408-3895
USA

and

G. Cornelis van Kooten

FEPA Research Unit
University of British Columbia
Vancouver, BC
Canada

The authors wish to acknowledge research support from the Sustainable Forest Management Network Centre of Excellence located at the University of Alberta, Edmonton, Canada.

ABSTRACT

Systematic biases away from true welfare estimates derived using dichotomous choice contingent valuation methodology (DCCV) are often thought to arise from flaws in survey design and delivery, preference uncertainty, or incorrect specification of functional form in the empirical model. This study examines the issue of accuracy of DCCV welfare estimates using a computational approach in which heterogeneous artificial agents with theoretically valid utility functions and full information are queried regarding their willingness to pay for an environmental good. We find that, even within an artificially created "perfect world," welfare estimates differ significantly from "true" welfare for agents with heterogeneous preferences.

INTRODUCTION

Increasing population and per capita consumption are cited as contributing factors to environmental degradation, including loss of biodiversity, deforestation and global climatic change (see Ehrlich and Holdren 1974; Vitousek et al. 1986; Krupa and Kickert 1989; Turner et al. 1990; Folke et al. 1996; Gowdy 1997). The ongoing evolution of complex adaptive ecological and economic systems is also fundamentally unpredictable (Arthur 1989; Nowak and May 1992) and threshold effects that might be exacerbated by human impact are to be expected (Bak and Chen 1991; Arrow et al. 1995). Given the potential magnitude of the value of nonmarket environmental amenities (Costanza et al. 1997), they cannot be ignored in policy making.

If cost-benefit analysis (CBA) is embraced as a decision-making methodology, then nonmarket valuation becomes critically important in the calculation of social benefits. Ignoring nonmarket values can lead to the significant underestimation of the economic benefits of conservation, a bias towards development in the decision-making process, and reduced social well being (Hausman 1993). The contingent valuation methodology (CVM) has become the principal means of valuing environmental goods, with CVM-derived values now used in CBA and in the assessment of damages in litigation under the 1980 Comprehensive Environmental Response, Compensation, and Liability Act (see Hausman 1993; Smith 1993; Arrow et al. 1995).

In this paper, we focus on a technical issue involving the effects of agent heterogeneity on the accuracy of welfare estimates derived using CVM. Economic analyses at the aggregate level typically make the simplifying assumption that economic agents are homogeneous and that, as a result, individual welfare measures, as represented by willingness to pay (WTP), can be aggregated to derive social costs and benefits of alternative policies or projects. Kirman (1992) argues that the "... reduction of the behaviour of a group of heterogeneous agents *even if they are all themselves utility maximizers*, is not simply an analytical convenience as often explained, but is both unjustified and leads to conclusions which are usually misleading and often wrong" (p.117, emphasis in original).

If errors generated from a CVM survey of a heterogeneous population are relatively minor or systematic, they might safely be ignored. However, if the bias in welfare estimates derived using CVM is relatively large and/or non-systematic, then their use in guiding environmental policy comes into question. But it is difficult in practice to assess the accuracy of nonmarket values derived using CVM in real-world situations (see, e.g., Arrow et al. 1993; Neill et al. 1994; Sagoff 1994; Cummings et al. 1997). In this study, we use an artificial market and computational approach to examine whether CVM provides valid estimates of aggregate social welfare when agents hold heterogeneous preferences with regards to environmental amenities. In particular, our null hypothesis is the following: Are social welfare measurements derived using CVM in an artificially-generated population of agents, who are heterogeneous in their degree of "environmental altruism," equal to the "true" social welfare measures for the population?

We use a direct utility function developed by Madriaga and McConnell (1987) and calculate associated "true" WTP (as measured by compensating surplus) of artificially-generated economic agents for changes in an environmental amenity. These true WTP values are then compared to WTP estimates derived using a dichotomous choice, contingent valuation device in the "perfect world" where there are no inconsistencies in survey design or delivery, where agents have valid utility functions, and where there is no agent preference uncertainty.

We begin in the next section by briefly reviewing the random utility maximization model that is used to estimate Hicksian welfare measures from dichotomous choice CVM questions. Then, in section 3, we outline the computational model that we use to conduct a dichotomous choice, CVM survey in a market of artificial agents. Our analysis of the artificial survey results is presented section 4, while discussion and implications regarding the validity of CVM are presented in the concluding section.

A REVIEW OF CVMs RANDOM UTILITY MAXIMIZATION MODEL

Compensating variation and equivalent variation are the theoretically correct Hicksian measures of consumer welfare to be used in CBA (Johansson 1993).[1] Suppose that the consumer has the following indirect utility function:

$$V = v(p, m, z) = U(x_i(p, m, z), z), \qquad [1]$$

where x is a vector of n market goods, p is a vector of associated prices, z denotes the quality and/or quantity of the environmental amenity, and m is household income. Assuming that the indirect utility function has the required properties from duality theory (see, e.g., Johansson 1993), one can analyze the effects of a shift in the consumer preference function due to changes in the level of the environmental amenity and derive theoretically-correct values of consumer

[1]Technically, the correct measures are compensating surplus and equivalent surplus.

welfare. Arrow et al. (1993) recommend that compensating variation be used as the welfare measure in CVM surveys.

The standard single-bound dichotomous choice contingent valuation (DCCV) method is based on Hanemann's (1984) Random Utility Maximization (RUM) model. With RUM, economic agents are asked to provide "yes" or "no" responses to a proposed "bid" (what they would have to pay) for designated changes in the quantity or quality of the environmental amenity. The DCCV methodology assumes that the responding agent knows his or her own utility function, which we denote as $U(i, m; s)$, where i is a binary choice variable (1 if the respondent is willing to pay the bid amount, 0 otherwise), m is income (as before) and s is a vector of other agent characteristics. To an independent observer, indirect utility can be modeled as a random variable (Park et al. 1991):

$$U(i, m; s) = V(i, m; s) + \varepsilon_{i,} \qquad [2]$$

which consists of a parametric probability distribution with mean $V(i,m;s)$ and random error that is independently and identically distributed (iid).

When faced with a bid, A, for a proposed change in the environmental amenity, the respondent will accept the bid if

$$v(1, m\text{-}A; s) + \varepsilon_1 > v(0, m; s) + \varepsilon_0. \qquad [2]$$

Relation [2] can be restated in terms of a probability as:

$$\Pr\{Yes\} = \Pr\{\, v(1, m\text{-}A; s) + \varepsilon_1 > v(0, m; s) + \varepsilon_0\} = F_\varepsilon\,(\Delta v), \qquad [3]$$

where $F_\varepsilon\,(\Delta v)$ represents the cumulative density function (cdf) of the respondent's true maximum WTP. This is commonly modelled as a logistic function in DCCV studies:

$$\Pr\{Yes\} = 1/(1 + e^{-\Delta V}). \qquad [4]$$

Two commonly used measures of welfare are the mean WTP and median WTP. The median is the fee for which the probability of a "yes" vote for a given level of change in the environmental amenity is 50%. The mean WTP can be calculated as the total area under the estimated cdf. Utility differences for both the linear and log-linear functions can easily be derived and used in the probabilistic framework to calculate median WTP:

Linear Model: Median WTP = $-\alpha/\beta \equiv w^*_{lin}$ [5]

Log-linear Model: Median WTP = $-\alpha\,(m/\beta) \equiv w^*_{log}$ [6]

Mean WTP values can be calculated by numerical integration of the area under the cdf curve:

Linear Model: $$\mathrm{E(WTP)} = \int_0^\infty [1–\mathrm{F}(b)]\,\mathrm{d}b \equiv \overline{w}$$ [7]

Log-linear Model: $$\mathrm{E(WTP)} \cong (1/\text{-}\beta)\ln(1+e^{\alpha}) \equiv \tilde{w}$$ [8]

If WTP is assumed to be non-negative, then mean WTP can be calculated from [7] or estimated using Hanemann's (1989) closed-form approximation of mean WTP for the log-linear utility function when that particular functional form is postulated [8]. $F(b)$ is the cumulative probability of a "no" response to the DCCV question and is a function of the bid amount A.

Mean WTP is preferred over median WTP as a true measure of consumer welfare on theoretical grounds because of desirable aggregation characteristics (Johansson et al. 1989). A problem can occur when using mean WTP if the bid curve does not converge (Cooper and Loomis 1992). Truncation and normalization of the mean is possible (Boyle and Bishop 1987), but the summation of individually-truncated welfare measures may provide a biased estimate of total consumer surplus due to the arbitrary nature of the truncation point.

Estimation of logit model coefficients is accomplished using the maximum likelihood method and leads to estimators that are asymptotically normal and have desirable asymptotic properties (Amemiya 1981). Parameter estimates calculated from CVM survey results, themselves random variables, can then be used to further calculate the non-linear random variable welfare measures (Bockstael and Strand 1987; Park et al. 1991). Approximate distributions of the value of estimated mean WTP can be derived from the RUM framework by applying bootstrapping techniques. In this regard, Park et al. (1991) develop WTP confidence intervals using the methods of Krinsky and Robb (1986). The technique is computationally intensive requiring at least 1,000 drawings of parameters for the development of confidence intervals for mean WTP and to account for all parameter interactions. In Monte Carlo studies, it is also possible to use a ranking and drop the upper and lower extreme values to derive confidence intervals using the 'percentile' method (Efron 1987; Dorfman et al. 1990)

AN ARTIFICIAL MARKET FOR APPLYING CVM

In this section, we develop a general procedure for generating an artificial population of economic agents that can be queried about their willingness to pay for changes in an environmental amenity. The goal is to choose parameters for the model in such a way that the distribution of true WTP follows a believable pattern for samples of artificial agents. The methodology of section 2 is used is used to analyze the responses of the artificial agents and develop estimated welfare measures that can be compared with the true welfare measures obtained directly from the agents' utility functions.

The utility function that we use is from Madriaga and McConnell (1987), who model a public environmental good that has both a pure existence value component and an influence on market goods through a weak complementary link. The direct utility function is:

$$U(x_1, x_2, R) = ax_1 + \ln(x_2) + b, \qquad [9]$$

where x_1 is a complementary environmental quality-related market good; x_2 is a strongly separable composite market good; $p_1x_1 + p_2x_2 = m$; and $p_2 = 1$. In this model, parameters a and b control the relative importance of market versus nonmarket goods in the utility function. The parameters are defined using reference and critical baseline levels for the environmental amenity:

$$a = \alpha\,(R - R_m), \text{ and} \qquad [10]$$

$$b = \exp(\beta\,(R - R_m)), \qquad [11]$$

where R is the level of the resource and R_m is a critical lower maintenance level of the resource. Parameter α impacts the size of use value in the utility function, while parameter β controls the contribution of the environmental good to existence value separate from the market good.

The RUM model utility difference framework can be used to calculate an agent's true WTP (Huang and Smith 1998):

$$\text{WTP}_i = \frac{(R_1 - R_0)m_i}{(R_1 - R_m)} + \frac{p_1}{\alpha_i(R_1 - R_m)}\left[\ln\left(\frac{R_0 - R_m}{R_1 - R_m}\right) + e^{\beta_i(R_1 - R_m)} - e^{\beta_i(R_0 - R_m)}\right]. \qquad [12]$$

Each agent's WTP is expressed as a function of income, price of the complementary market good, and the target and reference levels of the environmental amenity. This value can be calculated each time an agent is presented with a proposed combination of fee and level of environmental change, with a "yes" or "no" response generated based on comparing this WTP value with the proposed fee.

We experimented with a number of different parameterizations. The goal was to develop a model that showed WTP patterns typical of empirical studies where there is a rightwards skew with a very few respondents willing to accept high bids (Cooper and Loomis 1992).[2] The baseline parameters are the critical resource level, R_0=10.0; maintenance resource level, R_m=15.0; target resource level, R_1=16.5 (a 10% improvement in environmental quality); price of the weak complement, p_1= 0.5; and price of the composite good, p_2= 1.0.

The income parameter was normally distributed with mean 10.0 and standard deviation of 10%, $m \sim N(10.0, 1.0)$. The technical link parameter, α, was normally distributed with mean of 0.05 and standard deviation of 10%, $\alpha \sim N(0.050, 0.005)$. Finally, the β parameter, which defines the magnitude of environmental altruism for an agent in the model, was normally distributed with mean 0.125 and standard deviation 20% for non-altruists, $\beta_{NA} \sim N(0.125, 0.025)$, and with mean 0.225 and standard deviation 20% for environmental altruists, $\beta_A \sim N(0.225, 0.045)$. It should be noted that the ratio of WTP to income in the experiment is large; however, if household spending on charities, donations and environmental protection is assumed to be strongly separable from other household spending, the magnitude of the ratio is irrelevant.

One thousand samples, each a mix of 200 altruist and 200 non-altruists, were drawn from a heterogeneous population in which values of α, m and β varied and were normally distributed around appropriate means. Thus, a total of 400,000 artificial survey responses were generated (400 agents, 1,000 samples) and welfare measures calculated for them using each of the two assumed functional forms. The estimated welfare levels were then compared to the "true" welfare level using ANOVA.

Each sample, then, has a unique WTP distribution based on the randomly generated and normally distributed parameters for the model. The resulting distribution of WTP for Sample 87 is shown in Figure 1. There are two points to note: (1) the rightward skew of WTP is typical of the distributions seen in empirical studies; and (2) there are no zero WTP values. These are direct results of our choice of utility function and model parameters.

---

[2]No logit regressions were undertaken until the final decisions were made on model form and parameter values.

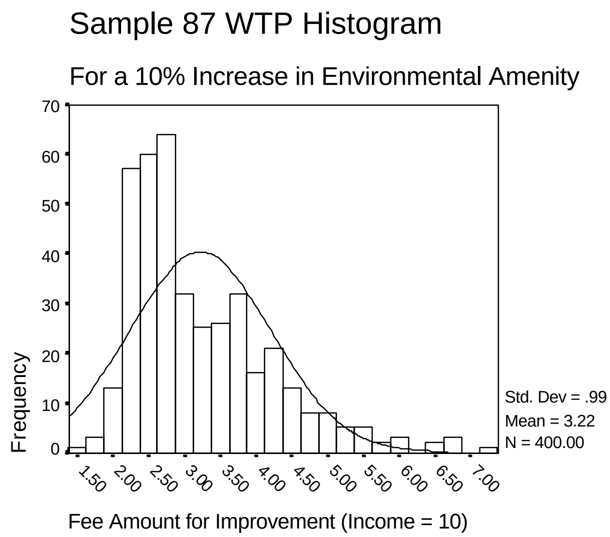

Figure 1: Willingness to Pay for an Increase in Environmental Quality

There are two basic schools of thought on bid vector design for DCCV studies. On the one hand, Boyle and Bishop (1987), Cooper and Loomis (1992), and Elnagheeb and Jordan (1995) advocate schemes in which there are a large number of fairly closely spaced bids over a wide range. A second perspective on bid vector design is advocated by Alberini (1995) and is followed in this study. Alberini found that maximum statistical efficiency for single-bounded, DCCV surveys occurred with only two bid values. There was very little gain in using designs with more than six to ten total bid points, or by placing bid values far out in the tails of the WTP distribution (i.e., where the probability of a "yes" response is less than 3%).

For each sample in this experiment, a total of six bid values were generated for presentation in the bid vector. The values of the generated bid vector were based on assumed normality of the WTP distribution (e.g., see the normal curve overlay of the WTP distribution shown in Figure 1). Conceptually, the normal approximation of the skewed distribution could be viewed as analogous to having the experimenter conduct focus-group research prior to finding an approximate range over which the bid amounts could be distributed.

We set the upper and lower bids at the points where 5% of the area under the normal curve was in the tails (i.e., $z = \pm 1.65$). The remaining bid values were calculated by dividing the normal distribution into even partitions:

$$\text{Division} = (B_U - B_L)/(N\text{-}1) = 0.18,$$

where B_U is the upper bid limit (0.95), B_L is the lower bid limit (0.05) and N is the total number of desired bids. All bids were rounded to the nearest 0.05. The distribution of bid partitions within the normal distribution is shown in Figure 2. Each agent in a sample was randomly presented with one of six potential bid amounts (extra low, very low, low, high, very high, extra high) and their response, based on their true WTP, was recorded.

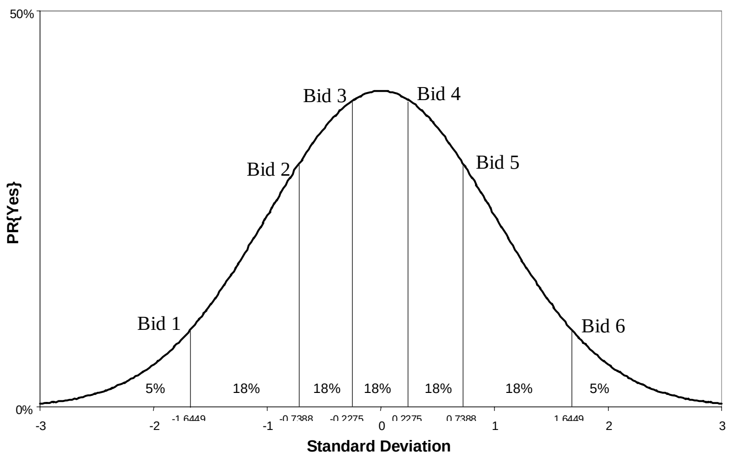

Figure 2: Determination of Bid Amounts

The 400 responses were then used in a logit regression, the coefficients of which could subsequently be used to construct welfare measures for each sample. Each logit regression was performed twice using SHAZAM (1993), once each for assumed linear and log-linear functional forms for utility. These functions are the ones assumed in empirical analysis and, clearly, not the true utility functions of the artificial agents. We calculated the median WTP for both the linear and log-linear utility function specifications, w^*_{lin} and w^*_{log}, respectively. We also calculated the mean WTP by numerical integration over three separate ranges from zero to: (1) the upper bid amount ($\overline{w}_{lin}^{100\%}$ and $\overline{w}_{\log}^{100\%}$); (2) 125% of the upper bid amount ($\overline{w}_{lin}^{125\%}$ and $\overline{w}_{\log}^{125\%}$); and (3) 150% of the upper bid amount ($\overline{w}_{lin}^{150\%}$ and $\overline{w}_{\log}^{150\%}$). The approximation of mean WTP, $\tilde{w}$, was also calculated and all welfare measures compared to true average utility, $\overline{w}_{true}$.

HOW ACCURATE ARE CVM MEASURES WHEN AGENTS HAVE HETEROGENEOUS PREFERENCES

Table 1 presents the regression results and calculated welfare measures for a 10% increase in the quality of the artificial environmental amenity. The table provides means and medians for both linear and log-linear utility functions, as well as true WTP and Hanneman's (1989) approximation of mean for the log-linear model. The true mean WTP distribution for the 1,000 samples is fairly tightly clustered around a mean of 3.261. Estimated WTPs are quite similar: the median WTP was 3.148 (s.e.=0.066) for linear utility and 3.156 (s.e.=0.066) for the log-linear utility difference, while numerical integration resulted in average values of mean WTP ranging from fc.126 to 3.159. True WTP, $\overline{w}_{true}$, was 3.2% to 4.2% higher than the various welfare estimates.

Table 1. Summary of Welfare Measures and Statistics

		Linear Functional Form				Log-Linear Functional Form				
	$\overline{w}_{true}$	w^*_{lin}	$\overline{w}^{100\%}_{lin}$	$\overline{w}^{125\%}_{lin}$	$\overline{w}^{150\%}_{lin}$	$w^*_{\log}$	$\overline{w}^{100\%}_{\log}$	$\overline{w}^{125\%}_{\log}$	$\overline{w}^{150\%}_{\log}$	$\tilde{w}_{log}$
Summary Statistics (N=1,000)										
Minimum	3.145	2.976	2.952	2.976	2.978	2.971	2.948	2.971	2.974	2.975
Mean	3.261	3.148	3.126	3.148	3.150	3.156	3.134	3.156	3.159	3.159
Maximum	3.363	3.348	3.331	3.348	3.350	3.361	3.348	3.361	3.361	3.362
Deviation	0.036	0.066	0.063	0.065	0.066	0.066	0.063	0.066	0.066	0.066
Skew	-0.015	0.122	0.115	0.128	0.137	0.141	0.130	0.144	0.148	0.150
Kurtosis	0.129	-0.184	-0.093	-0.161	-0.163	-0.072	0.103	-0.060	-0.087	-0.086
Confidence Intervals										
Using ± 1.96 s.e										
Lower C.I.	3.190	3.019	3.002	3.019	3.021	3.027	3.010	3.028	3.030	3.030
Upper C.I.	3.333	3.276	3.251	3.276	3.279	3.286	3.258	3.285	3.288	3.289
Using Percentile										
Lower C.I.	3.190	3.021	3.002	3.022	3.025	3.027	3.006	3.028	3.031	3.032
Upper C.I.	3.330	3.276	3.252	3.276	3.285	3.281	3.254	3.280	3.283	3.284

The values of mean WTP at 125% and 150% truncation levels are virtually identical indicating that the bid curve does converge. Positive skew and variable kurtosis statistics are indicative of non-normality for all distributions. The second part of Table 1 shows 95% confidence intervals calculated using standard errors and by percentile methods.

The distributions of true, mean and median WTPs for the 1,000 samples in the experiment are shown in Table 2. With 1,000 samples, welfare estimates are asymptotically normally distributed and it is therefore possible to undertake simple single-factor ANOVA and test the null hypotheses that various welfare measures are equal. Table 3 shows the results of these analyses for a variety of groupings. The null hypothesis of interest is listed in the left-hand column.

Table 2. Distribution of True and Calculated Welfare Measures for 1,000 Samples

		Linear Functional Form				Log-Linear Functional Form				
WTP Range	$\overline{w}_{true}$	w^*_{lin}	$\overline{w}^{100\%}_{lin}$	$\overline{w}^{125\%}_{lin}$	$\overline{w}^{150\%}_{lin}$	$w^*_{\log}$	$\overline{w}^{100\%}_{\log}$	$\overline{w}^{125\%}_{\log}$	$\overline{w}^{150\%}_{\log}$	$\tilde{w}_{log}$
< 2.950	0	0	0	0	0	0	1	0	0	0
2.951 to 2.975	0	0	6	0	0	1	5	1	1	1
2.976 to 3.000	0	11	17	11	11	9	9	9	9	8
3.001 to 3.025	0	18	14	17	17	9	18	9	7	7
3.026 to 3.050	0	25	73	26	21	20	70	20	16	17
3.051 to 3.075	0	85	108	86	80	88	88	92	90	84
3.076 to 3.100	0	102	143	105	110	76	93	76	75	81
3.101 to 3.125	0	133	149	138	123	102	169	92	93	92
3.126 to 3.150	1	159	156	149	159	170	169	176	172	168
3.151 to 3.175	15	134	115	140	138	157	118	159	152	157
3.176 to 3.200	19	124	70	118	124	121	108	126	128	122
3.201 to 3.225	140	57	84	70	65	88	69	81	79	85
3.226 to 3.250	205	82	39	70	80	61	47	70	77	77
3.251 to 3.275	272	43	18	42	43	62	28	53	65	64
3.276 to 3.300	213	19	4	19	20	26	1	26	23	24
3.301 to 3.325	84	3	0	4	3	2	3	2	5	5
3.326 to 3.350	36	5	4	5	6	2	4	4	2	2
3.351 to 3.375	15	0	0	0	0	6	0	4	6	6
> 3.376	0	0	0	0	0	0	0	0	0	0

For the case of the linear utility function, the null hypotheses of equality between the true welfare measure and each of the separate estimated welfare measures are strongly rejected. The hypotheses that all four welfare estimates, the median and the mean calculated at three different truncation levels, and true WTP, are equal are also strongly rejected. The results are essentially the same for the model that uses a log-linear functional form. In that case, the F-statistics are lower than in the linear utility model but still highly statistically significant.

The results of ANOVA indicate that the hypotheses of equality of median welfare estimates and mean welfare estimates calculated using truncation levels of 125% and 150% of the upper bid level cannot be rejected. For the log-linear utility specification, the approximation of mean (equation [4]) also provides a welfare estimate that cannot be distinguished from median or mean welfare estimates calculated using truncation levels of 125% and 150% of the upper bid level.

Table 3. ANOVA Results

Null Hypothesis	Degrees Freedom	Linear Form	Log-Linear Form
Intra-Form Hypotheses			
$w^* = \overline{w}^{100\%} = \overline{w}^{125\%} = \overline{w}^{150\%} = \overline{w}_{true}$	4 and 4,995	793.8***	693.0***
$w^* = \overline{w}^{100\%} = \overline{w}^{125\%} = \overline{w}^{150\%}$	3 and 3,996	29.2***	33.2***
$\overline{w}^{100\%} = \overline{w}^{125\%} = \overline{w}^{150\%}$	2 and 2,997	40.5***	46.0***
$\overline{w}^{100\%} = \overline{w}^{125\%}$	1 and 1,998	54.1***	61.6***
$\overline{w}^{100\%} = \overline{w}^{150\%}$	1 and 1,998	67.0***	77.3***
$\overline{w}^{125\%} = \overline{w}^{150\%}$	1 and 1,998	0.8	0.9
$w^* = \overline{w}_{true}$	1 and 1,998	2300.5***	1944.1***
$\overline{w}^{100\%} = \overline{w}_{true}$	1 and 1,998	3411.2***	3090.5***
$\overline{w}^{125\%} = \overline{w}_{true}$	1 and 1,998	2313.0***	1975.5***
$\overline{w}^{150\%} = \overline{w}_{true}$	1 and 1,998	2187.0***	1852.8***
$\tilde{w}_{log} = w^*$	1 and 1,998	n/a	1.04
$\tilde{w}_{log} = \overline{w}^{100\%}$	1 and 1,998	n/a	80.7***
$\tilde{w}_{log} = \overline{w}^{125\%}$	1 and 1,998	n/a	1.31
$\tilde{w}_{log} = \overline{w}^{150\%}$	1 and 1,998	n/a	0.04
$w^* = \overline{w}^{100\%}$	1 and 1,998	54.5***	63.1***
$w^* = \overline{w}^{125\%}$	1 and 1,998	0.001	0.014
$w^* = \overline{w}^{150\%}$	1 and 1,998	0.75	0.69
Inter-Form Hypotheses			
$w^*_{lin} = w^*_{log}$	1 and 1,998	8.97***	
$\overline{w}^{100\%}_{lin} = \overline{w}^{100\%}_{log}$	1 and 1,998	6.48**	
$\overline{w}^{125\%}_{lin} = \overline{w}^{125\%}_{log}$	1 and 1,998	8.54***	
$\overline{w}^{150\%}_{lin} = \overline{w}^{150\%}_{log}$	1 and 1,998	8.74***	

True WTP for the heterogeneous population of environmental altruists and non-altruists is significantly different from the WTP estimates derived using the DCCV methodology. It appears that neither of the empirical models can adequately account for the relatively simple utility model developed by Madriaga and McConnell. The utility function is additive and non-linear in the level of environmental change; the empirical models do not fully account for the non-market values held by heterogeneous economic agents.

The last section of Table 3 provides F-statistics for tests of the equivalence of welfare measures obtained using linear and log-linear utility specifications. For median and mean welfare estimates calculated at all three truncation levels, the null hypothesis of equality of measures derived using linear and log-linear utility functional forms was rejected at the 5% or 1% levels of statistical significance. Significant differences in welfare estimates do result from using alternative functional forms.

DISCUSSION

The main goal of this study was to address the technical question of whether estimated welfare measures derived using CVM could provide accurate approximations of true welfare values when agent preferences for nonmarket environmental amenities are heterogeneous. Using a computational method, we reject hypotheses that true and estimated welfare measures are equal in an artificial population of economic agents who possess well-defined utility functions and who are heterogeneous in their degree of environmental altruism. In addition, the DCCV methodology used in this study resulted in the rejection of hypotheses of the equality of welfare measures derived using linear and log-linear functional forms.

Although the statistical significance of these results is not an issue, there are likely to be different views as to the seriousness of the implications. The true welfare measure, compensating surplus, for the artificial population was only 3% to 4% higher than the welfare estimates derived using the DCCV methodology. This margin of error is not particularly high, especially when compared to potential sources of error in “real-world” studies. Further, concern about WTP estimates derived from CVM studies has been that they were too high rather than too low (Huang and Smith 1998), but we found the opposite—in our case, the estimated welfare measures actually underestimated the true welfare impacts.

On the other hand, the welfare estimates in this study were derived within a “perfect world.” The myriad of potential problems with CVM surveys are eliminated when we use computational experiments and artificial agents and markets. It is not unreasonable, therefore, to expect that tools used to measure well being should at least be accurate in this perfect world. If DCCV estimates are not accurate under conditions of perfect information and rationality, should the tool be trusted to provide data that is used in real-world policy decisions?

This is a normative question which this study cannot answer definitively although our results suggest that estimated welfare measures are unreliable when there is agent heterogeneity in environmental preferences, as is the case in the "real world." This occurs even under simple conditions when preferences are specified in accordance with economic theory, agents have complete information, and there are no technical complications relating to survey design and delivery. Thus, even if one accepts all the necessary assumptions required to use DCCV results in CBA, however debatable they may be, there is still no guarantee of their accuracy.

Possibilities for calibration of DCCV welfare estimates might still exist, but further investigation is needed. If DCCV estimates are consistently lower than true welfare for populations with heterogeneous environmental preferences then calibration might be possible. There does not, however, seem to be any strong *a priori* reasons why this should be the case. Ideally, this study design could be expanded and used on a variety of functional forms for direct and indirect utility to shed more light on this question, as well as different assumptions regarding the distribution of true WTP.

REFERECES

Alberini, A. (1995). Testing willingness-to-pay models of discrete choice contingent valuation survey data. *Land Economics* 71: 83–95.

Amemiya, T. (1981). Qualitative response models. *Journal of Economic Literature* 9: 1483–1536.

Arrow, K., B. Bolin, R. Costanza, P. Dasgupta, C. Folke, C. S. Holling, B.-O. Jansson, S. Levin, K.-G. Mäler, C. Perrings and D. Pimentel (1995). Economic growth, carrying capacity, and the environment. *Science* 268: 520–521.

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner and H. Schuman (1993). Advance notice of proposed rulemaking, extension of comment period and release of contingent valuation methodology report. *Federal Register* 58: 4601–4614.

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal* 99:116–131

Bak, P. and K. Chen (1991). Self-organized criticality. *Scientific American* (January, 1991): 46–53.

Bockstael, N. and I. E. Strand Jr. (1987). The effect of common sources of regression error on benefit estimates. *Land Economics* 63: 11–20.

Boyle, K. J. and R. C. Bishop (1987). Valuing wildlife in benefit-cost analyses: a case study involving endangered species. *Water Resources Research* 23: 943–950.

Cooper, J. and J. Loomis (1992). Sensitivity of willingness-to-pay estimates to bid design in dichotomous choice contingent valuation models. *Land Economics* 68: 211–224.

Costanza, R., R. d'Arge, R. de Groot, S. Farber, M. Grasso, B. Hannon, K. Limburg, S. Naeem, R. V. O'Neill, J. Paruelo, R. G. Raskin, P. Sutton and M. van den Belt (1997). The value of the world's ecosystem services and natural capital. *Nature* 387: 253–260.

Cummings, R. G., S. Elliot, G. W. Harrison and J. Murphy (1997). Are hypothetical referenda incentive compatible? *Journal of Political Economy* 105: 609–621.

Dorfman, J., C. Kling and R. Sexton (1987). Confidence intervals for elasticities and flexibilities: re-evaluating the ratio of normals case. American Journal of Agricultural Economics 72: 1006–1017.

Efron, B. (1990). Better bootstrap confidence intervals. Journal of the American Statistical Association 82: 171–185.

Ehrlich, P. R. and J. P. Holdren (1974). Impact of population growth. *Science* 171: 1212–1217.

Elnagheeb, A. H. and J. L. Jordan (1995). Comparing three approaches that generate bids for the referendum contingent valuation method. *Journal of Environmental Economics and Management* 29: 92–104.

Folke, C., C. S. Holling and C. Perrings (1996). Biological diversity, ecosystems, and the human scale. *Ecological Applications* 6: 1018–1024.

Gowdy, J. (1997). The value of biodiversity: markets, society, and ecosystems. *Land Economics* 73: 25–41.

Hanemann, W. M. (1984). Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics* 66: 332–341.

Hanemann, W. M. (1989). Welfare evaluations in contingent valuation with discrete response data: response. *American Journal of Agricultural Economics* 71: 1057–1061.

Hausman, J. A. (1993). *Contingent Valuation: A Critical Assessment*. Elsevier.

Huang, J. C. and V. K. Smith (1998). Monte Carlo benchmarks for discrete response valuation methods. *Land Economics* 74: 186–202.

Johansson, P.-O. (1993). *Cost-benefit Analysis of Environmental Change*. Cambridge University Press.

Johansson, P.-O., B. Kriström and K. G. Mäler (1989). Welfare evaluations in contingent valuation with discrete response data: comment. *American Journal of Agricultural Economics* 71: 1054–1056.

Kirman, A. P. (1992). Whom or what does the representative individual represent? *Journal of Economic Perspectives* 6: 117–136.

Krinsky, I. and A. L. Robb (1986). On approximating the statistical properties of elasticities. *Review of Economics and Statistics* 68: 715–719.

Krupa, S. V. and R. N. Kickert (1989). The greenhouse effect: impacts of ultraviolet-B (UV-B), radiation, carbon dioxide (CO2) and ozone (O3) on vegetation. *Environmental Pollution* 61: 263–393.

Madriaga, B. and K. E. McConnell (1987). Exploring existence value. *Water Resources Research* 23: 936–942.

Neill, H., R. G. Cummings, P. Ganderton, G. Harrison and T. McGlucking (1994). Hypothetical surveys and real economic commitments. *Land Economics* 70: 145–154.

Nowak, M. A. and R. M. May (1992). Evolutionary games and spatial chaos. *Nature* 359: 826–829.

Park, T., J. Loomis and M. Creel (1991). Confidence intervals for evaluating benefits estimates from dichotomous choice contingent valuation studies. *Land Economics* 67: 64–73.

Sagoff, M. (1994). Should preferences count? *Land Economics* 70: 127–144.

SHAZAM (1993). *SHAZAM User's Reference Manual Version 7.0*. McGraw-Hill Book Company.

Smith, V. K. (1993). Nonmarket valuation of environmental resources: an interpretative appraisal. *Land Economics* 69: 1–26.

Turner, B. L., W. C. Clark, R. W. Kates, J. F. Richards, J. T. Mathews and W. B. Meyers (1990). *The Earth as Transformed by Human Action: Global and Regional Changes in the Biosphere Over the Past 300 Years*. Cambridge University Press.

Vitousek, P. M., P. R. Ehrlich, A. Ehrlich and P. Matson (1986). Human appropriation of the products of photosynthesis. *BioScience* 36: 368–373.