Journal of Environmental Economics and Policy, 2013
http://dx.doi.org/10.1080/21606544.2013.775602

UKNEE
UK Network of Environmental Economists

Estimating willingness to accept using paired comparison choice experiments: tests of robustness

David C. Kingsley[a*] and Thomas C. Brown[b]

[a]*Department of Economics, University of Massachusetts Lowell, 1 University Ave. Lowell, MA 01854, USA;* [b]*Rocky Mountain Research Station, U.S. Forest Service, 240 W. Prospect Road, Fort Collins, CO 80526, USA*

(*Received 7 November 2012; final version received 7 February 2013*)

Paired comparison (PC) choice experiments offer researchers and policy-makers an alternative nonmarket valuation method particularly apt when a ranking of the public's priorities across policy alternatives is paramount. Similar to contingent valuation, PC choice experiments estimate the total value associated with a specific environmental good or service. Similar to choice experiments, the questions posed to respondents are choices between alternatives. In contrast to both methods, respondents in PC choice experiments make choices between pairs of dissimilar alternatives including private goods, public goods, and monetary amounts. The alternatives may include competing policy alternatives, thus providing a ranking of the public's priorities among those alternatives. We investigate the robustness of estimated welfare measures to econometric modelling and choice set composition across two PC choice experiments. Results suggest that accounting for repeated observations increases the efficiency of welfare estimates but also reveals, contrary to previous research, sensitivity to choice set composition. Thus, while PC choice experiments may be advantageous in certain situations the results presented here suggest that further research is needed to better understand the sensitivities of the resulting welfare estimates.

Keywords: nonmarket valuation; willingness to accept; paired comparison choice experiments; environmental policy

JEL Code: Q58, Q51

1. Introduction

Although public goods and services are not priced within markets, they may nonetheless be of substantial value to society. Governments face difficult choices in deciding which public goods and services to provide and in what quantity. Such choices are aided by an estimate of the value the public places on these goods and services. The contingent valuation method (CVM) and discrete choice experiments (DCEs) have emerged as the primary stated preference methods for estimating the economic value of environmental and other nonmarket goods and services (Kanninen 2007; Mitchell and Carson 1989). CVM presents respondents with a single specified good (or service) and, using one format or another, asks respondents about their willingness to pay (WTP) for that good. In discrete choice CVM, the distribution of individual WTP is identified by varying the cost of the good across respondents. On the other hand, a DCE presents respondents with choices between two or more alternatives that typically vary along two dimensions, first in levels of attributes of the environmental good (the commodity dimension) and second in cost

*Corresponding author. Email: David_Kingsley@uml.edu

(the price dimension) (Day and Prades 2010). Respondents' choices are used to estimate individual WTP for changes in the attribute levels of the nonmarket good.

The method of paired comparisons (PCs) is an alternative stated preference method that, like a DCE, relies on choices between alternatives. The PC method dates back to Fechner (1860), and the fundamental characteristics of the method have been studied extensively (e.g., Bock and Jones 1968; David 1988; Thurstone 1927; Torgerson 1958). The PC method obtains an individual respondent's preference order among the various items of a choice set by presenting the items in pairs and asking respondents to choose the item in each pair that best meets a given criterion, such as being the more preferred. When used for economic valuation the items include dollar amounts along with several dissimilar goods, possibly a mix of public and private goods (Peterson and Brown 1998). By forcing choices between goods, the PC method helps to make the tradeoffs between goods explicit and to remind subjects of policy alternatives. The choices between private goods and money are familiar, providing consumers with a benchmark from which to make the more difficult choices between two public goods or between a public good and a monetary amount. Further, when the choice set includes multiple public goods, the method provides a ranking of those goods. Research suggests that when respondents make choices between alternative gains, PC choice experiments provide a conservative measure of willingness to accept (WTA) (i.e., one unaffected by loss aversion) (Champ and Loomis 1998; Loomis et al. 1998; Peterson and Brown 1998).

Three related approaches have been used to estimate welfare measures from PC data. Peterson and Brown (1998) calculate an individual's WTA for each good by bracketing goods between dollar amounts based on the items' preference scores. An item's preference score is simply the number of times a respondent chose the item over the other items in the choice set. Because all items in the choice set, both goods and dollar amounts, receive preference scores, each good is easily placed between two different dollar amounts. In a similar analysis Champ and Loomis (1998) bracket the value of a good by the smallest dollar amount chosen over the good and the largest dollar amount rejected in favour of the good. Within the bracket, they then estimate individual WTA using a weighted midpoint estimated from the individual preference scores. Champ and Loomis (1998) found the mean WTA inferred from PC data to be robust to context and scope effects. Specifically they found no statistical difference in measured WTA for a public good in two distinctly different choice sets. However, there was a significant difference in measured WTA when the scope of the public good was altered. Loomis et al. (1998) further developed the method by introducing a parametric model that controls for important covariates and is fashioned after the Welsh and Bishop (1993) approach for analysing payment card CVM data. The Loomis et al. (1998) research suggested that the WTA estimated from the PC choice experiment provided a WTA-WTP value disparity that, while still significant, was lesser in magnitude than if measured using conventional methods, supporting the notion that PC yields a conservative estimate of individual WTA.

This paper investigates the robustness of welfare measures estimated from data of two PC choice experiments. The first of two objectives is to develop a random effects (RE) probit model designed for PC choice experiments that accounts for the repeated observations in the data, and compare the performance of the RE model to that of an interval model similar to that used by Loomis et al. (1998). Results indicate that accounting for repeated observations increases the efficiency (lower standard errors) of both the estimated coefficients and the mean WTA. This finding is consistent with that of a similar analysis performed on payment card data (Alberini, Boyle, and Welsh 2003). Further, results

suggest that these differences are significant for several of the goods included in each of the PC data-sets.

The second objective is to test the sensitivity of welfare estimates to the composition of the PC choice set. The choice sets of the two PC experiments examined here include six goods in common plus four goods unique to each choice set. We produced welfare estimates for each set using the interval and RE models. Using the interval model, we found, as did Champ and Loomis (1998), that welfare measures are robust to choice set composition. However, using the RE model reveals a cautionary note: the estimated values may be more sensitive to the bundle of goods and monetary amounts included in the PC choice set than previously suggested. Using the RE model, the differences in mean WTA are significant for four of the six common goods.

2. Data and methods

We utilise data from two prior PC experiments. The first data-set, referred to as Treatment 1, was originally analysed by Peterson and Brown (1998); the other, Treatment 2, was analysed by Rosenberger et al. (2003). In both data-sets all items were gains and the respondents were instructed to choose the item in each pair they would prefer if they could have either at no cost. Subjects for each treatment were recruited from undergraduate courses at Colorado State University and independently participated in scheduled sessions in a computer lab on campus. Table 1 displays the number of participants and descriptive statistics collected within each treatment.

The Treatment 1 choice set consisted of four private goods (a $15 Meal, a $200 Clothing Certificate, $75 Tickets and Transportation to a cultural or sporting event, and a $500 certificate good for Airline Tickets), six locally relevant public goods (a Wildlife Refuge, a Clean Air and Water Agreement, a Spring Festival, a free library Video Service, an increase in campus Parking Capacity, and an increase in the Eating Area on campus) (Table 2), and 11 monetary amounts (Table 3). Goods were not paired with themselves and dollar amounts were not compared. Each respondent made 155 choices, 45 between goods and 110 between a good and a dollar amount. For example, one of the choices was between the Wildlife Refuge and $100, and another choice was between the Wildlife Refuge and Parking Capacity. For presentation, the order in which the 155 pairs were presented was separately randomised for each respondent. Of 330 respondents, four were dropped because of missing data, leaving 326 respondents providing a total of 50,530 individual observations of which 35,860 are between a good and a dollar amount.

The Treatment 2 choice set included similar monetary amounts to those of Treatment 1 (Table 3) but changed some of the goods (Table 4). Six goods were common across the two treatments, namely Clothing Certificate, Videotape Service, Parking Capacity,

Table 1. Descriptive statistics by treatment[a].

	Treatment 1	Treatment 2
N	326	123
Age	19.85	19.45
Gender	0.55	0.6
Education	13.18	13.31

[a]Mean age in years; mean gender (0 = female, 1 = male); mean years of education.

Table 2. Goods in Treatment 1 data.

Private goods
1. A meal at a Fort Collins restaurant of your choice, not to exceed $15. (**$15 Meal**)
2. Two tickets and transportation to *one* of the following: A Colorado ski area of your choice. A concert of your choice in Denver (Contemporary or Classical). A Broncos, Rockies, or Nuggets game. A cultural event of your choice at the Denver Center for the Performing Arts. Estimated value: $75. (**$75 Tickets**)
3. A nontransferable $200 certificate for clothing at a Fort Collins store of your choice. (**$200 Clothing Certificate**)
4. A nontransferable certificate for you to make $500 worth of flights on an airline of your choice. (**$500 Travel Certificate**)
Public goods
1. A no-fee library service that provides videotapes of all course lectures so that students can watch tapes of lectures for classes they are not able to attend. (**Videotape Service**)
2. Parking garages to increase parking capacity on campus such that students are able to find a parking place at any time, without waiting, within a five-minute walk of any building at no increase in the existing parking permit fee. (**Parking Capacity**)
3. Purchase by CSU of 2000 acres of land in the mountains west of Fort Collins as a wildlife refuge for animals native to Colorado. (**Wildlife Refuge**)
4. A CSU-sponsored, on-campus springtime weekend festival with a variety of live music and student participation events with no admission fee. (**Spring Festival**)
5. Expansion of the eating area in the Lory Student Center to ensure that any student can find a seat at any time. (**Eating Area**)
6. A cooperative arrangement between CSU, local business groups, and the citizens of the community that would ensure the air and water of Fort Collins would be at least as clean as the cleanest 1% of the communities in the US. (**Clean Air and Water**)

Wildlife Refuge, Eating Area, and Clean Air and Water, while the remaining goods differed (Tables 2 and 4). The 123 respondents each made 145 choices, 100 choices between goods and dollar amounts and 45 between goods, for a total of 17,835 observations of which 12,300 are between a good and a dollar amount.

Table 3. Monetary amounts in Treatments 1 and 2.

Treatment 1	Treatment 2
1	10
25	30
50	50
75	70
100	100
200	200
300	300
400	400
500	500
600	700
700	

Table 4. Goods in Treatment 2 data.

Private goods
1. A meal at a Fort Collins fast food or cafeteria-style restaurant of your choice. (**Meal**)
2. A $25 gift certificate for use at a bookstore of your choice. (**$25 Book Certificate**)
3. Two tickets to ONE of the following: (**Tickets**) A concert of your choice in Denver (contemporary, new age, hard rock, or classical) General admission to a professional sport event in Denver (Broncos, Rockies, Nuggets, or Avalanche)
4. A nontransferable $200 certificate for clothing at a Fort Collins store of your choice. (**$200 Clothing Certificate**)
5. A nontransferable certificate for a round-trip flight to any major city in the contiguous 48 states on an airline of your choice. (**Travel Certificate**)
Public goods
1. A no-fee library service that provides videotapes of all course lectures so that students can watch tapes of lectures for classes they are not able to attend. (**Videotape Service**)
2. Parking garages to increase parking capacity on campus such that students are able to find a parking place at any time, without waiting, within a five-minute walk of any building at no increase in the existing parking permit fee. (**Parking Capacity**)
3. Purchase by CSU of 2000 acres of land in the mountains west of Fort Collins as a wildlife refuge for animals native to Colorado. (**Wildlife Refuge**)
4. Expansion of the eating area in the Lory Student Center to ensure that any student can find a seat at any time. (**Eating Area**)
5. A cooperative arrangement between CSU, local business groups, and the citizens of the community that would ensure the air and water of Fort Collins would be at least as clean as the cleanest 1% of the communities in the US. (**Clean Air and Water**)

Two covariates, Age and Gender, were recorded for all respondents. In addition, in Treatment 2, the variable Environment was recorded, indicating whether the respondent belonged to at least one environmental advocacy organisation.

To provide a comparison of the interval model to the RE model we used the approach developed by Loomis et al. (1998) to create data-sets including only those choices that maintain consistency between a good and the presented monetary amounts. Recall that each respondent chooses between each good and a series of dollar amounts. If the respondent is consistent, the WTA associated with the good will be bracketed by the largest dollar amount over which the individual chooses the good (this is the individual's lower bound on WTA) and the smallest dollar amount chosen over the good (this is the individual's upper bound on WTA). To illustrate, consider a series of four choices between good *Z* and dollar amounts $5, $10, $15, and $20. If good *Z* is an economic gain we might expect the individual to choose good *Z* over either $5 or $10 but to choose both $15 and $20 over good *Z*. Therefore, $10 is the largest dollar amount over which the individual chooses the good, establishing the lower bound, and $15 is the smallest dollar amount chosen over the good, establishing the upper bound. In this case the respondent's WTA lies in the range [10, 15]. For those respondents who display choice inconsistency (e.g., choose the good over $5 and $15 but choose $10 and $20 over the good), their choices between the good and dollar amounts were removed from the sample (only the choices between the specific good and dollar amounts were removed, not the entire set of choices made by that respondent). This procedure establishes a consistent subset of choices each subject made between a good and a dollar amount, producing 24,772 (from a possible 35,860) observations in Treatment 1 and 9570 (from a possible 12,300) in Treatment 2.[1]

3. Econometric specification

PC data provide discrete observations on choice data that are readily analysed. We first specify the underlying WTA function and then derive the two likelihood functions required for the interval and RE models. The notation is as follows: individuals are represented by i, choice occasion by j, and goods by k. Choice occasion is simply the sequence of choices (1st choice, 2nd choice, etc.) an individual encounters. Monetary amounts, represented by t_{ij}, provide the bid level for respondent i on choice occasion j. Thus,

$$w_{ijk} = \alpha_k + \beta_k X_i + \varepsilon_{ijk}, \tag{1}$$

where w_{ijk} represents the unobserved WTA associated with good k for respondent i on choice occasion j, α_k represents a good-specific constant, β_k represents the good-specific coefficients for each included covariate (X_i), and $\varepsilon_{ijk} \sim N(0, \sigma^2_{\varepsilon_k})$. The restriction on the error term follows directly from the data used. Recall that only the subsets of consistent choices within each good are used; thus, there is a different set of individuals and choice occasions across each good.[2] By taking advantage of the variation in t_{ij}, both the location, the mean WTA for each good, and the scale, the standard deviation of the WTA distribution, are identified (Cameron 1988; Cameron and James 1987; Cameron and Quiggin 1994).

The likelihood function for the interval model, as proposed by Loomis et al. (1998), defines t_{i_l} as respondent i's lower bound on WTA (equal to the largest dollar amount over which the item is chosen) and t_{i_u} as i's upper bound on WTA (equal to the smallest dollar amount chosen over the item). The probability statement $P_i(t_{i_l})$ gives the probability that the individual's WTA is less than t_{i_l}, so

$$P_i(t_{i_l}) = Pr(u_{ijk} < t_{i_l}) = Pr(\alpha_k + \beta_k X_i + \varepsilon_{ijk} < t_{i_l}) = \Phi\left(\frac{(t_{i_l} - \alpha_k - \beta_k X_i)}{\sigma_{\varepsilon_k}}\right) \tag{2}$$

and we can define the likelihood function as:

$$\boldsymbol{L}(\boldsymbol{\alpha_k}, \boldsymbol{\beta_k}, \boldsymbol{\sigma_{\varepsilon_k}}) = \prod_{\boldsymbol{i}} \left(\boldsymbol{P_i}(\boldsymbol{t_{i_u}}) - \boldsymbol{P_i}(\boldsymbol{t_{i_l}})\right). \tag{3}$$

The specification for the RE probit model treats each choice as a draw from the individual's WTA distribution. To account for the repeated observations within individual this model allows these observations to be correlated but maintains the assumption of independence across individuals. In this model $P_i(t_{ij})$ is the probability that the individual's WTA is less than t_{ij}, so

$$\boldsymbol{P_i}(\boldsymbol{t_{ij}}) = \boldsymbol{Pr}(\boldsymbol{\alpha_k} + \boldsymbol{\beta_k X_i} + \boldsymbol{\varepsilon_{ijk}} + \boldsymbol{g_i} < \boldsymbol{t_{ij}}) = \Phi\left(\frac{\boldsymbol{t_{ij}} - \boldsymbol{\alpha_k} - \boldsymbol{\beta_k X_i} - \boldsymbol{g_i}}{\boldsymbol{\sigma_{\varepsilon_k}}}\right). \tag{4}$$

The additional term $g_i \sim N(0, \sigma_i)$, as compared with the specification for the interval model, represents the RE. The dependent variable, y_{ijk}, equals 1 if the monetary amount is chosen over the good and 0 otherwise. The likelihood function can be

constructed as follows:

$$L(y_{ijk};\alpha_k,\beta_k,\sigma_{\varepsilon_k}) = \prod_i \prod_j \left(1 - P_i(t_{ij})\right)^{1-y_{ijk}} P_i(t_{ij})^{y_{ijk}}. \quad (5)$$

4. Results

We first investigate the impact of accounting for the repeated observations in the data using the RE model. Recall that the interval approach reduces each individual's sequence of choices to a single interval. Our primary interests are the magnitude and standard errors of the estimated coefficients and the means and standard errors of the estimated WTAs. Results suggest that using the RE model has a substantial effect on both the coefficients and the estimated WTAs. We then turn to the question of whether the composition of the PC choice set influences WTA. Results suggest that using the RE model reveals, contrary to previous research, that the composition of the PC choice set has a significant effect on mean WTA.

4.1. Econometric modelling

Table 5 displays the results of the interval and RE probit models estimated using Treatment 1 data. The numbers of observations (N) per good in each model, listed in Table 5, depend on the number of consistent observations within each good and indicate the subset of respondents, from among the full set of 326 respondents, whose choices between the good and each dollar amount were consistent. Across goods, roughly 70% of the subjects provided consistent choices.[3] For the RE model, Table 5 also lists the coefficient of correlation between individual responses, ρ, for each good.[4] Results indicate that for both covariates (Age and Gender) the estimated standard errors are substantially lower using the RE model as opposed to the interval model, resulting in greater significance of the included covariates. Specifically, the RE model suggests that Age has a significant impact on WTA of four goods, two more (Wildlife Refuge and the Eating Area) than with the interval model. Similarly, using the RE model Gender is significantly related to WTA of six goods, two more (Wildlife Refuge and the Spring Festival) than with the interval model.

Beyond a reduction in the standard errors, accounting for the repeated observations of each subject has a considerable impact on the magnitude of the estimated coefficients (Table 5). To be succinct, we focus our attention on those coefficients that are determined to be significant in both the interval and RE models. For Age, the impact of choosing the RE model over the interval model is slight for the Travel Certificate (the coefficients are −$7.71 with the interval model and −$7.42 with the RE model) but is substantial for the Clothing Certificate (−$4.51 versus $4.55). Similarly for Gender the impact of model selection is slight for the Clothing Certificate ($32.15 versus $33.80) and the Travel Certificate ($105.88 versus $93.15) but is substantial for the Eating Area ($28.46 versus $11.95).

Using the variation in t_{ij} to identify the WTA function allows us to interpret the estimated coefficients analogously to ordinary least squares (Cameron 1988; Cameron and James 1987; Cameron and Quiggin 1994). That is, each coefficient indicates the estimated change in WTA, measured in dollar terms, for a one-unit change in the associated covariate. For example, consider the coefficient on Age, measured in years; the interval model suggests that a one-year increase in a respondent's age *decreases* individual WTA of the Clothing Certificate by $4.51, whereas the RE model suggests an additional year *increases* individual WTA by $4.55.

Table 5. Results for the interval and random effects models, Treatment 1 data[a].

	Interval					Random effects					
Good	*N*	Log L	Intercept	Age	Gender	*N*	Log L	ρ	Intercept	Age	Gender
Videotape Service	211	−656	111.01	1.76	51.99	2321	−722	0.88	197.27	0.77	14.23
			(112.39)	(5.54)	(36.37)				(27.03)***	(1.25)	(12.91)
Parking Capacity	227	−663	195.62	−3.61	43.18	2497	−750	0.96	25.15	0.65	−6.95
			(108.46)	(5.32)	(34.28)				(13.95)	(0.68)	(5.05)
$15 Meal	265	−302	28.81	−0.54	1.24	2915	−263	0.96	21.94	−0.29	0.41
			(7.18)***	(0.35)	(2.32)				(4.48)***	(0.22)	(1.34)
$200 Clothing Certificate	235	−442	241.47	−4.51	32.15	2585	−482	0.90	68.79	4.55	33.80
			(44.75)***	(2.19)**	(14.88)**				(15.62)***	(0.75)***	(7.42)***
Wildlife Refuge	224	−508	601.62	−1.56	69.62	2464	−599	0.96	191.58	10.29	37.08
			(221.19)***	(10.69)	(75.82)				(30.58)***	(1.52)***	(9.64)***
Spring Festival	225	−686	67.39	2.76	−14.96	2475	−717	0.96	47.47	−0.87	8.65
			(86.77)	(4.25)	(27.17)				(11.57)***	(0.57)	(3.70)**
$75 Tickets	215	−549	35.78	3.57	22.61	2365	−558	0.98	65.45	0.34	4.72
			(47.87)	(2.33)	(16.58)				(9.54)***	(0.47)	(3.15)
Eating Area	222	−515	−27.17	3.55	28.46	2442	−551	0.91	−2.94	1.36	11.95
			(43.14)	(2.10)	(13.53)**				(10.96)	(0.53)**	(3.74)***
$500 Travel Certificate	211	−432	521.84	−7.71	105.88	2321	−448	0.90	524.91	−7.42	93.15
			(74.72)***	(3.64)**	(26.67)***				(41.84)***	(1.97)***	(24.16)***
Clean Air and Water	217	−573	17.94	17.28	151.04	2387	−630	0.95	423.08	0.85	54.03
			(220.39)	(10.86)	(62.71)**				(45.14)***	(2.07)	(13.63)***

[a]Standard errors in parentheses.
** (***) indicates the coefficient is significant at the $p < 0.05$ ($p < 0.01$) level.

Table 6. Estimated willingness to accept by treatment[a].

	Treatment 1			Treatment 2		
Good	Interval	RE	Difference	Interval	RE	Difference
Videotape Service	198.02	226.90	−28.88	226.11	74.06	152.05
	(49.79)***	(12.79)***	(51.41)	(59.45)***	(7.69)***	(59.95)**
Parking Capacity	167.24	31.03	136.21	315.09	155.23	159.86
	(23.03)***	(3.65)***	(23.32)***	(74.19)***	(16.44)***	(75.99)**
$200 Clothing Certificate	184.02	192.95	−8.93	167.26	167.43	−0.17
	(10.05)***	(5.22)***	(11.33)	(13.06)***	(13.01)***	(18.43)
Wildlife Refuge	640.16	433.71	206.45	621.75	510.35	111.40
	(53.77)***	(8.44)***	(54.43)***	(106.36)***	(31.68)***	(110.98)
Clean Air and Water	513.03	494.05	18.98	636.80	414.39	222.41
	(45.53)***	(17.19)***	(48.67)	(104.03)***	(32.21)***	(108.9)**
Eating Area	71.77	36.03	35.74	66.78	53.84	12.94
	(8.97)***	(2.49)***	(9.31)***	(5.54)***	(12.27)***	(13.46)
$75 Tickets	129.42	76.97	52.45			
	(11.09)***	(2.18)***	(11.30)***			
Tickets				128.09	108.27	19.82
				(23.12)***	(5.47)***	(23.76)
$500 Travel Certificate	473.37	469.51	3.86			
	(18.22)***	(17.45)***	(25.23)			
Travel Certificate				345.84	326.75	19.09
				(39.88)***	(13.99)***	(42.26)
$15 Meal	19.41	16.55	2.86			
	(1.55)***	(0.95)***	(1.82)			
Meal				10.77	10.79	−0.02
				−21.20	(5.41)**	(21.88)
Spring Festival	107.26	38.73	68.54			
	(18.72)***	(2.56)***	(18.89)***			
$25 Book Certificate				27.86	27.85	0.01
				(5.43)***	(2.75)***	(6.09)

[a]Goods included in both treatments are listed first. The difference is calculated within treatment (interval WTA – RE WTA). WTAs are estimated at mean Age, Gender = 1 (Male) and, in Treatment 2, Environment = 1 (belonging to at least 1 environmental advocacy group). The standard errors of the estimated WTA, and differences, are approximated using the Delta Method and are reported in parentheses.
** (***) indicates the coefficient is significant at the $p < 0.05$ ($p < 0.01$) level.

The differences observed in the coefficients translate to differences in the mean and standard error of the estimated WTAs. Mean WTA is estimated using the coefficients of Table 5 with Age set at its mean and Gender set to 1 (Male). The standard errors of the estimated WTAs are approximated using the Delta Method.[5] As is the case with the coefficients, the standard error associated with each mean WTA is lower using the RE model (e.g., for the Videotape Service, the standard errors are 50 with the interval model and 13 with the RE model) (Table 6). Within treatment the significance of the difference (interval WTA – RE WTA) is again found using the Delta Method and indicates that accounting for repeated observations (by using the RE model) significantly alters the WTA of five of the 10 goods (Table 6), as follows: Parking Capacity, $167 with the interval model versus $31 with the RE model, for a difference of $136 (82%); Wildlife Refuge, a difference of $207 (32%); Eating Area, $36 (50%); $75 Tickets, $53 (41%); and Spring Festival, $69 (64%). The effect in each case is to lower the estimated WTA. Four of these five

goods are public goods, the values of which would be of interest to policy-makers. Further, the estimated WTA of the $75 Tickets ($129 with the interval model but $77 with the RE model) becomes more reflective of the expressed monetary value. While we cannot definitely state that the RE model produces WTA estimates that more accurately reflect underlying preferences, these results, together with the increased efficiency, are encouraging.[6]

Analysis of the Treatment 2 data yields similar results. The lower standard errors associated with the coefficients in the RE model, compared with the interval model, affect the implied significance of the covariates (Table 7). Regarding the magnitudes of the covariates, and focusing again on the change in value of the coefficients that are significant in both models, we find that the coefficients on Gender are similar across models for the $200 Clothing Certificate ($24.14 with the interval model versus $22.60 with the RE model) and for the $25 Book Certificate ($7.18 versus $7. 17). However, for the parameter Environment we observe substantial differences across the two models for Clean Air and Water ($363 versus $156) and Videotape Service ($137 versus $34). Interpretation of this coefficient implies, based on the RE model results, that WTA for the Clean Air and Water good would be, on average, $156 *higher* for individuals who belong to at least one environmental advocacy organisation than for individuals who do not belong to such organisations.

As with the Treatment 1 data, these differences translate to differences in the mean and standard error of the estimated WTAs. To estimate WTA for the Treatment 2 data, Age is set to its mean, Gender to 1, and Environment to 1 (indicating affiliation with at least one environmental advocacy group). The estimated mean WTAs significantly differ between model for three of the 10 goods (Table 6). In each case, using the RE model results in a lower mean WTA; for the Videotape Service the reduction is $152 (67%), for the Parking Capacity it is $160 (51%), and for the Clean Air and Water it is $222 (35%). Again, these differences are nontrivial and suggest a lower estimated WTA for at least a subset of the included public goods.

4.2. Choice set composition

Finally, we consider the estimated WTAs across treatment within model for the set of six common goods. In each case the difference is calculated as Treatment 1 WTA – Treatment 2 WTA (presented in Table 6) and the standard error of the difference is calculated using the Delta Method. Consistent with previous research, the WTAs of the common goods appear robust to the composition of the PC choice set using the interval model, as none of the reported differences are significant (Table 8). However, comparing the WTAs estimated using the RE model leads to a different conclusion. This analysis finds that the values of four of the six goods are significantly different across treatment (Table 8), as follows: Video Service has an estimated value of $227 in Treatment 1 and $74 in Treatment 2, for a difference of $153 or 67%; Parking Capacity, a difference of −$124 (400%); Wildlife Refuge, −$77 (18%); and Clean Air and Water, $80 (16%). Note that there is no systematic difference in estimated WTA across treatment, implying that the choice set composition affects each item differently.[7] It also should be noted that the choice sets differed in both the included goods and dollar amounts. We suspect that the dollar amounts are sufficiently similar across choice sets (Table 3) as to suggest that the observed differences are driven by the differences among the included goods, but we cannot unambiguously decompose the relative impacts on WTA.

Table 7. Results for the interval and random effects models, Treatment 2 data[a].

	Interval						Random effects						
Good	*N*	Log L	Intercept	Age	Gender	Environment	*N*	Log L	ρ	Intercept	Age	Gender	Environment
Parking Capacity	84	−235	142.96	−2.69	104.47	120.55	840	−252	0.96	−4.57	8.37	42.02	−46.52
			(174.85)	(8.02)	(69.84)	(76.79)				(26.03)	(1.22)***	(14.85)***	(15.12)***
Tickets	100	−235	153.99	−2.15	−10.14	26.27	1000	−240	0.97	43.57	2.38	22.78	−4.68
			(55.74)***	(2.58)	(22.71)	(24.63)				(13.34)***	(0.60)***	(6.01)***	(5.52)
$200 Clothing Certificate	101	−96	139.65	0.13	24.12	1.06	1010	−98	0.86	138.06	0.35	22.60	−0.07
			(25.53)***	(1.21)	(11.98)**	(13.56)				(17.77)***	(0.99)	(11.17)**	(13.25)
Eating Area	96	−217	−30.77	0.61	19.13	66.48	960	−228	0.95	54.18	−3.33	50.58	14.77
			(110.35)	(5.12)	(39.87)	(45.78)				(23.12)**	(1.09)***	(10.16)***	(12.06)
$25 Book Certificate	115	−117	12.20	0.24	7.18	3.78	1150	−235	0.89	12.20	0.24	7.17	3.77
			(6.75)	(0.31)	(2.67)***	(2.95)				(6.79)**	(0.32)	(2.68)***	(2.96)
Clean Air and Water	83	−227	59.57	5.30	110.45	362.82	840	−247	0.96	220.12	0.00	38.12	156.14
			(232.36)	(10.95)	(90.81)	(107.20)***				(56.41)***	(2.53)	(19.71)**	(29.26)***
Videotape Service	89	−243	−116.36	8.56	39.56	136.91	890	−260	0.98	23.59	0.67	3.63	33.84
			(137.69)	(6.68)	(56.17)	(63.50)**				(18.33)	(0.96)	(7.47)	(8.18)***
Wildlife Refuge	86	−215	363.19	−4.02	144.64	193.32	870	−242	0.95	123.14	14.25	73.10	33.01
			(284.96)	(12.94)	(110.32)	(116.08)				(72.81)	3.41***	(24.29)***	(23.12)
Travel Certificate	90	−174	439.17	−7.61	19.64	35.69	900	−184	0.95	146.29	7.06	17.74	24.73
			(85.63)***	(3.99)	(36.39)	(41.89)				(29.77)***	(1.53)***	(13.25)	(15.75)
Meal	113	−123	5.22	0.03	2.06	2.93	1130	−246	0.94	5.22	0.03	2.15	2.87
			(11.24)	(0.52)	(4.92)	(5.49)				(11.80)	(0.55)	(5.01)	(5.73)

[a]Standard errors in parentheses.
** (***) indicates the coefficient is significant at the $p < 0.05$ ($p < 0.01$) level.

Table 8. Difference in estimated willingness to accept across treatment[a].

Good	Interval	RE
Videotape Service	−28.09	152.84
	(77.55)	(14.93)***
Parking Capacity	−147.85	−124.21
	(77.68)	(16.84)***
$200 Clothing Certificate	16.76	25.52
	(16.48)	(14.02)
Wildlife Refuge	18.41	−76.64
	(119.18)	(32.78)***
Clean Air and Water	−123.77	79.66
	(113.56)	(36.51)***
Eating Area	4.98	−17.81
	(10.54)	(12.52)

[a]The difference is calculated across treatment (Treatment 1–Treatment 2). The standard error of this difference, in parentheses, is estimated using the Delta Method.
*** indicates the coefficient is significant at the $p < 0.01$ level.

5. Conclusion

This research has two primary results. Firstly, we investigate the impact of using a RE probit model that accounts for the repeated observations within subject in two PC choice experiments. This model is compared to an interval model that reduces each subject's set of choices between a good and a sequence of dollar amounts to a single interval within which the WTA resides. Results suggest that accounting for the repeated observations increases the efficiency of the welfare measure and in some cases results in significant differences in estimated WTA across model. Specifically, five significant differences were observed in the Treatment 1 data and three were observed in the Treatment 2 data. Further, the effect was always to lower WTA, with the reductions ranging from 32% to 82% and averaging 52%. Although no conclusion can definitely be made regarding which set of welfare estimates better reflects underlying preferences, the greater efficiency of the RE model argues in its favour.

Secondly, the analysis suggests that, in contrast to the interval model, the RE model reveals that the estimated WTAs of the goods in PC experiments are sensitive to the composition of the choice set. That is, the estimated WTAs of the included public goods and services appear to be influenced by the set of private goods and monetary amounts in the choice set.

In conclusion, given the greater efficiency of the RE model and the encouraging measures of WTA obtained, this research suggests that the RE model may better reflect underlying preferences within pair comparison choice experiments. However, this research also confirms the concern that choice set composition within a PC experiment may influence estimates. In the presence context effects, great care must be undertaken to ensure a neutral effect so that estimated WTAs are not biased in one direction. As with other stated preference methods, the values measured using the PC method are sensitive to the econometric specification and to rather subtle changes in the decision context. Nevertheless, it remains that PC choice experiments offer policy-makers and researchers an alternative nonmarket valuation technique with important features. The PC method uniquely provides a WTA welfare measure across a variety of policy alternatives. Especially when policy alternatives are being compared, a PC choice experiment offers a sensible

alternative to CVM or DCEs. Additional research is warranted to better understand the limitations and sensitivities exposed here.

Acknowledgements

The authors would like to thank Patricia Champ, Nicholas Flores, and Donald Waldman as well as seminar participations at the University of Massachusetts Amherst for helpful comments. All errors remain our own.

Notes

1. This procedure is necessary to estimate the interval model proposed by Loomis et al. (1998) and is used with within the RE model for comparability.
2. We do not address whether preferences change over the sequence of choices here. Recent research has investigated the intransitivity of preferences, preference learning, and the structure of the error process in PC (Brown et al. 2008; Brown and Peterson 2009; Kingsley and Brown 2010). Results from the Treatment 1 data suggest that the variance of the error process decreases in a way consistent with preference learning. Note that had fatigue been present the variance of the error process would have been expected to increase over the sequence of choices.
3. Peterson and Brown (1998) investigate the interaction of choice consistency and preferences in the Treatment 1 data.
4. The coefficient of correlation between responses from the same individual, ρ, is quite close to 1 across all goods reflecting the choice consistency in the data and indicating the presence of RE.
5. As is commonly applied, the Delta Method is used to calculate the square root of $\bar{x}' Cov(\hat{\beta})\bar{x}$ where $Cov(\hat{\beta})$ is the covariance matrix of the estimated coefficients and $\bar{x}$ represents the mean of the included covariates. Note that the significance or insignificance of individual item intercepts or coefficients does not alter the interpretation of estimated WTA and the respective standard errors (Cameron and James 1987; Cameron 1988; Cameron and Quiggin 1994).
6. For comparison, mean WTAs for the Treatment 1 goods from Peterson and Brown (1998), which they estimated from subjects' preference scores, are as follows: Videotape Service ($198), Parking Capacity ($171), $200 Clothing Certificate ($187), Wildlife Refuge ($394), Clean Air and Water ($328), Eating Area ($95), $75 Tickets ($154), $500 Travel Certificate ($388), $15 Meal ($48), and Spring Festival ($154). Rosenberger et al. (2003) did not report estimated WTAs; their research explores measures of environmental disposition on the marginal effects of each good.
7. The lack of a systematic difference across Treatments is not surprising given that these experiments were not explicitly design to explore choice set composition effects. However, this finding suggests that choice set composition can significantly affect estimated mean WTA and thus a careful investigation to better understand the effect is warranted.

References

Alberini, A., K. Boyle, and M. Welsh. 2003. "Analysis of Contingent Valuation Data with Multiple Bids and Response Options Allowing Respondents to Express Uncertainty." *Journal of Environmental Economics and Management* 45 (1): 40–62.

Bock, R. D., and L. V. Jones. 1968. *The Measurement and Prediction of Judgment and Choice*. San Francisco, CA: Holden-Day, Inc. 370.

Brown, T. C., D. Kingsley, G. L. Peterson, N. Flores, A. Clarke, and A. Birjulin. 2008. "Reliability of Individual Valuations of Public and Private Goods: Response Time, Preference Learning, and Choice Consistency." *Journal of Public Economics* 92 (7): 1595–1606.

Brown, T. C., and G. L. Peterson. 2009. *An Enquiry into the Method of Paired Comparison: Reliability, Scaling, and Thurstone's Law of Comparative Judgment*. General Technical Report RMRS-GTR-216WWW. Fort Collins, CO: Rocky Mountain Research Station.

Cameron, T. A. 1988. "A New Paradigm for Valuing Non-Market Goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression." *Journal of Environmental Economics and Management* 15 (3): 355–379.

Cameron, T. A., and M. D. James. 1987. "Efficient Estimation Methods for "Closed-Ended" Contingent Valuation Surveys." *Review of Economics and Statistics* 69 (2): 269–276.

Cameron, T. A., and J. Quiggin. 1994. "Estimation Using Contingent Valuation Data from a "Dichotomous Choice with Follow-up" questionnaire." *Journal of Environmental Economics and Management* 27: 218–234.

Champ, P. A., and J. B. Loomis. 1998. "WTA Estimates Using the Method of Paired Comparison: Tests of Robustness." *Environmental and Resource Economics* 12 (3): 175–386.

David, H. A. 1988. *The Method of Paired Comparisons*. 2nd ed. New York: Oxford University Press. 188.

Day, B., and J.-L. P. Prades. 2010. "Ordering Anomalies in Choice Experiments." *Journal of Environmental Economics and Management* 59 (3): 271–285.

Fechner, G. T. 1860. *Elemente der Psychophysik*. Leipzip: Breitkopf and Hartel.

Kanninen, B. J., ed. 2007. *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice*. Dordrecht: Springer.

Kingsley, D. A., and T. C. Brown. 2010. "Preference Uncertainty, Preference Learning and Paired Comparison Choice Experiments." *Land Economics* 86 (3): 530–544.

Loomis, J., G. Peterson, P. Champ, T. Brown, and B. Lucero. 1998. "Paired Comparison Estimates of Willingness to Accept Versus Contingent Valuation Estimates of Willingness to Pay." *Journal of Economic Behavior and Organization* 35: 501–515.

Mitchell, R. C., and R. T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington, DC: Resources for the Future. 463.

Peterson, G. L., and T. C. Brown. 1998. "Economic Valuation by the Method of Paired Comparison, with Emphasis on Evaluation of the Transitivity Axiom." *Land Economics* 74 (2): 240–261.

Rosenberger, R. S., G. L. Peterson, A. Clarke, and T. C. Brown. 2003. "Measuring Dispositions for Lexicographic Preferences of Environmental Goods: Integrating Economics, Psychology and Ethics." *Ecological Economics* 44 (1): 63–76.

Thurstone, L. L. 1927. "A Law of Comparative Judgment." *Psychology Review* 34: 273–286.

Torgerson, W. S. 1958. *Theory and Methods of Scaling*. New York: John Wiley & Sons. 460.

Welsh, M. P., and R. C. Bishop. 1993. "Multiple Bounded Discrete Choice Models." W-133 Benefits and Costs in Natural Resource Planning, Sixth Interim Report. Department of Agricultural and Applied Economics, University of Georgia, Athens, GA.