

WTA Estimates Using the Method of Paired Comparison: Tests of Robutness

PATRICIA A. CHAMP¹ and JOHN B. LOOMIS²

¹USDA Forest Service, 3825 E. Mulberry Street, Fort Collins, CO 80524, USA; ²Department of Agricultural and Resource Economics, Colorado State University, Fort Collins, CO 80523, USA

Received 3 December 1997; accepted 19 September 1997

Abstract. The method of paired comparison is modified to allow choices between two alternative gains so as to estimate willingness to accept (WTA) without loss aversion. The robustness of WTA values for two public goods is tested with respect to sensitivity of the WTA measure to the context of the bundle of goods used in the paired comparison exercise and to the scope (scale) of the public good project. There was no statistical difference in WTA measures for open space based on independent treatments with different contexts. One treatment involved valuing open space within a set of goods with similar value. The other treatment involved valuing open space within a choice set of goods which had lower dollar values than open space. There was a statistical difference in WTA between a permanent expansion in the bus system and a temporary expansion in the bus system. We conclude the method of paired comparison appears to be a very promising approach to elicit measures of WTA.

Key words: context effect, public goods experiment, scope test, stated preferences, willingness to accept

1. Introduction

The need to measure the social benefits of environmental amenities is increasingly important. A variety of stated and revealed preference methods can be used to measure use values associated with a resource. However, measuring non-use values poses more of a challenge. Contingent valuation (CV) allows for measurement of non-use benefits in addition to use values. While there is some consensus that CV willingness to pay (WTP) estimates are valid (Carson et al. 1996), the same cannot be said of CV willingness to accept (WTA). The current practice in the United States, as suggested by the National Oceanic and Atmospheric Administration (1993), is to use WTP as a proxy for WTA when measuring welfare losses associated with an environmental degradation. WTA is clearly the appropriate welfare measure and as Bromley (1995: 129) points out “If the damages from pollution accidents are incorrectly measured, then the shadow price of spills and accidents will be wrong and therefore incorrect signals will be sent to both producers and consumers. These incorrect signals will result in too many spills and accidents”. This article discusses the results of an experiment in which an alternative approach to CV called the method of paired comparison (PC) is used to measure WTA.

Magat et al. (1988) applied the PC approach to non-market valuation as a means of measuring WTP. Peterson et al. (1996) extended this work by using PC to measure WTA. The PC approach provides a ranking of stated preferences for public goods and dollar amounts (or private goods and dollar amounts). From this ranking, WTA dollar values for the goods in the choice set are inferred. As with CV, PC can measure both use and non-use values. Magat et al. (1988: 396) suggest that one of the virtues of PC is the simplicity of the exercise relative to a CV exercise, "Because of the ease that subjects have in making paired comparisons, as opposed to more precise responses required in a hypothetical market, this methodology may offer a greater opportunity for eliciting individual preferences in contingent contexts". Furthermore, initial studies (Peterson et al. 1996; Peterson and Brown 1997; Loomis et al. 1996) suggest that the PC WTA estimates are consistent with basic tenets of economic theory. We assess the robustness of WTA estimates for public goods based on PC generated data. Specifically, two tests are conducted. As the method of PC has respondents value a good within the context of a bundle of goods, the first test examines the effect of the composition of the choice set on the value estimates for an open-space project. The second test compares the value estimates from independent samples for a permanent expansion of municipal bus service with a temporary expansion for a weekend for evidence of scope.

2. The Method of Paired Comparison

The method of paired comparison goes back to Fechner (1860) and has been the subject of much psychological research as a means to measure preferences (Thurston 1929; David 1988). The PC exercise in this study elicited participants' preferences for goods and sums of money via a series of binary choices between either two goods or a good and a sum of money. The choice sets which are the focus of this study contained four public goods and ten sums of money. The paired comparisons were made from the chooser reference point and the individual chooses between two gains.¹ For example, the individual chooses between getting \$50 or the provision of a public good. If she selects the public good, WTA for the public good is inferred to be greater than \$50. The chooser reference point may avoid loss aversion associated with the standard CV approach for measuring WTA (Franciosi et al. 1993; Kahneman et al. 1990).

PC is similar to contingent ranking and conjoint analysis in that it is an approach developed in a discipline outside economics for purposes other than non-market valuation. Whereas contingent ranking and conjoint analysis are traditionally used to predict choice or preferences as a function of attributes², we use PC to elicit choices between different goods or goods and dollar amounts with the ultimate goal of bracketing WTA. PC could be used as an elicitation method in a conjoint analysis where the individual chooses between goods which vary in attribute levels.

We are suggesting that PC could be used as an alternative non-market valuation technique to CV. There are at least two conceptual advantages of PC over the

traditional or single bound dichotomous choice CV. First, individuals value goods within the context of a bundle of goods. The number of goods in the bundle and the type of competing goods can be varied to mimic policy relevant trade-offs. If a public agency can provide only one or two public goods or services, the choice set can include the relevant trade-offs. Most CV surveys consider just one good, or at most three (Hoehn and Loomis 1993). Second, the repetitive choices between different dollar amounts and the good provide the opportunity to bracket the individual's WTA between a lower and upper dollar amount.

3. Test for a Context Effect

3.1. CONTEXT EFFECTS

Given that the PC approach involves choices within the context of a bundle of goods, one might expect different bundles to create contexts which affect the interpolated values of the goods. Parducci (1968) observed a context effect in an experiment in which college students were asked to rate the moral value of various behaviors. Each act was to be rated as 1 "not particularly bad or wrong", 2 "undesirable, a good person would not do this", 3 "wrong, highly questionable", 4 "seriously wrong", 5 "extremely evil". The experiment was designed such that two lists of acts were used with one act common to both lists. One of the lists contained acts which were "nasty" while the other list contained relatively mild acts. Students were explicitly instructed to treat each act independently. Despite this instruction, students tended to give the act common to both treatments a higher rating when it was embedded in the mild list relative to when it was embedded in the nasty list. Brown and Daniel (1987) also found evidence of a context effect when asking people to rate scenic beauty. They asked two independent samples to rate scenic quality of various scenes on a ten point scale with the endpoints labeled as 1 "very low scenic quality" and 10 "very high scenic quality". Treatment 1 was given a "low scenic beauty" set of scenes to rate while Treatment 2 was given a "high scenic beauty" set. Both treatments were also given a common set of scenes. The common scenes were presented in the same order in both treatments. As with the Parducci (1968) study, Brown and Daniel (1987) found evidence of a context effect. Specifically, they found that the scenes common to both treatments were rated significantly higher by the treatment group given the low scenic beauty set prior to the common set. If there is evidence of a context effect with the PC approach, the issue becomes one of deciding which context is most appropriate. This could be a rather subjective and controversial task.

Context effects have also been found in CV exercises. Randall et al. (1981) found that the sequence in which goods are valued in a CV exercise affects the value estimates. The estimated value of a public good tends to get smaller as the good moves down the valuation sequence (Randall and Hoehn 1992). PC attempts to avoid sequencing effects by randomizing the order of the choice sets.

Table I. Choice sets.

Variable	Description
<i>Choice Set for Treatment Group 1 (N=53):</i>	
BIK	Bike Trail. Connect the existing bike trails in the City of Fort Collins by adding a trail so as to make possible a continuous loop. The addition would allow one to go north and south on a completely dedicated bike path that would connect existing trails as shown on the map.
OS	Open Space. Purchase by the City of 125 acres of privately owned land and lakes along the Poudre River where the current bike trail goes. This property is adjacent to an identical piece of City owned open space. If the private land is <i>not</i> purchased it would be developed for residential housing and one-storey office buildings.
AQ	Air Quality Improvement. Smoke from old wood stoves, dust from sand left on roads long after snowstorms and tailpipe emissions from old automobiles often cause haze that obscures the view of the mountains on over 50 days a year (see photo A on board). This program would fund retrofitting of old wood burning stoves to improve combustion and reduce smoke particles, pay for city crews to sweep up sand on streets immediately after the snow has melted rather than waiting until spring and pay owners of old (pre-1975's) cars to retire their older polluting cars. Visibility would be improved to the level shown in photo B.
BUS	Expanded bus routes and time. Free city bus service to all residents of Fort Collins, additional routes and expanded hours of service. The additional routes are shown on the poster. The hours of operation would be expanded from the current 7:30 am to 6:30 pm to 6 am to midnight with buses running every 15 minutes instead of every half hour. The expansion of the city bus service would reduce traffic congestion and air pollution.
<i>Choice Set for Treatment Group 2 (N=50):</i>	
TRL	Perform 100 feet of trail improvements at Lory State Park. The program would bring in topsoil to repair 100 feet of erosion from heavy use on the Arthur's Rock hiking and mountain biking trail at Lory State Park. This would repair current erosion, slow future erosion and allow for natural revegetation of the 100 feet of trail.
OS	Open Space. Purchase by the City of 125 acres of privately owned land and lakes along the Poudre River where the current bike trail goes. This property is adjacent to an identical piece of City owned open space. If the private land is <i>not</i> purchased it would be developed for residential housing and one-storey office buildings.
RAP	Raptor Rehabilitation. One injured raptor such as a hawk or eagle under the care of the Rocky Mountain Center in Fort Collins would be given one dose of antibiotics. This dose would speed up the bird's recovery.
BSH	Extend time of bus service on Labor Day weekend. The hours of operation would be expanded <i>from</i> the current 7:30 am to 6:30 pm <i>to</i> run from 6 am to midnight on Saturday, Sunday and Monday of Labor Day weekend.

3.2. HYPOTHESIS

Using independent samples, we test whether value estimates based on PC data are susceptible to a context effect. We set up an experiment to test the equality of

value estimates using PC data from two different choice sets. The goods used in the two choice sets are described in Table I. The open space project was common to both choice sets. In choice set 1, we expected all goods to be of similar value; whereas in choice set 2, the value of open space was expected to be higher than the value of the other goods. Our null hypothesis ($WTA^{OS}_1 = WTA^{OS}_2$) was that the estimates of WTA for open space would not vary within the context of the bundle of goods. In other words, if the value estimates are robust we do not expect them to be sensitive to the value other goods in the choice set.

4. Tests for Sensitivity to Scope

4.1. SCOPE EFFECTS

The standard approach to CV has been criticized (Kahneman and Knetsch 1992; Desvousges et al. 1992) for its inability to provide value estimates for public goods that vary with the quantity or quality of the good.³ Kahneman and Knetsch (1992) argue that they find an embedding effect in which contingent values for a good vary depending on whether the good is valued by itself or as part of a more inclusive good. Desvousges et al. (1992) conducted an experiment in which they find estimates of total value for a public good do *not* increase for large increases in the level of the good provided. The conclusion of both these studies is that CV is not an appropriate method to be used for measuring natural resource damage assessments.⁴

4.2. HYPOTHESIS

If the values estimates based on PC data are consistent with the tenets of economic theory, they should be sensitive to scope. To test this hypothesis, we had independent samples participate in PC exercises with two versions of a public good that varied in scope. Specifically, choice set 1 included a good which described a program to permanently expand bus routes and the hours of a free bus service in the city (BUS) while choice set 2 included a good which described a program to temporarily extend the current hours of bus service for one weekend (BSH). Granted the difference in scope between the two goods of interest is quite substantial; however, the Desvousges et al. (1992) study suggests that contingent values are not even sensitive to substantial changes in the quantity of resources provided. Future experiments might test the sensitivity to more subtle differences. Table I provides a description of both goods. The null hypothesis is that the value of BUS is equal to BSH ($H_0: WTA^{BUS}_1 = WTA^{BSH}_2$) with the alternative being ($H_a: WTA^{BUS}_1 > WTA^{BSH}_2$).

4.3. STATISTICAL METHODS

We use a non-parametric approach to estimate mean and median WTA.⁵ This approach involves a weighted linear interpolation between the highest dollar amount over which the individual prefers the good and lowest dollar amount at which she prefers the money over the good. For example, an individual might prefer the open-space project over \$120 but prefers \$180 over the open-space project. Since the interval where she switched from preferring the open space to the money is between \$120 and \$180, we know WTA lies within this interval. Rather than simply using the mid-point we rely on additional choice information in the PC data to calculate where in this interval the individual's minimum WTA is likely to be. Specifically, we calculate a weighted mid-point using the number of times the individual chooses this good over the other goods in the choice set and over the different dollar amounts (i.e., the rank order of the good over the other goods and monetary amounts). In our example, if \$120 was chosen 6 times, the good was chosen over other goods 7 times and \$180 was chosen 9 times, the weighted interpolation would calculate WTA at \$140 rather than the pure midpoint of the interval (\$150) since the rank order of the good (7) is closer to the rank order of \$120 (6) than it is to \$180 (9). This process of weighted interpolation is repeated for each individual and a sample mean, median and standard deviation are calculated. For these experiments, the highest dollar amount included in the paired comparisons was \$295. If an individual preferred the good over the highest dollar amount, we inferred that her maximum WTA was \$300. Clearly, this approach is somewhat conservative but it does not affect the conclusions drawn from this experiment. However, the measures of central tendency will be very susceptible to the rule used to infer maximum WTA in cases where individuals prefer the good over the highest dollar value.

5. Data

5.1. SUBJECTS

University clerical and administrative staff in academic and non-academic units were recruited and paid \$20 for attending one of the 45 minute sessions held on campus. The sessions were conducted before work, at lunch and after work.

5.2. LABORATORY EXPERIMENTS

This experiment was conducted in a laboratory setting. Each study participant was seated at a computer terminal. The computer code presented the pairs of items (either two goods or one good and a dollar amount) on the monitor in random order for each respondent.⁶ For each pair, the respondent chose the preferred good or sum of money. Respondents were told "when the choice appears on the screen, please choose or vote for the one that you would like to receive. The choice selection by

Table II. Demographic characteristics of the study participants.

<i>Variable</i>	<i>Treatment 1 (N = 53)</i>	<i>Treatment 2 (N = 50)</i>
<i>Age</i>		
21–30 years of age	12%	30%
31–40 years of age	43%	26%
41–50 years of age	33%	26%
More than 50 years of age	12%	17%
<i>Gender</i>		
Male	13%	14%
Female	87%	86%
<i>Education Level</i>		
High school graduate	14%	16%
Some post high school education	33%	28%
College graduate	32%	28%
Some post college education	21%	28%

the majority would be provided to everyone. For example, if a majority chooses the dollar amount, it would be received by all households in the County and the public program would NOT be provided. If the public program is chosen by the majority, it would be available to you as well as anyone else in the County”.⁷

6. Summary Statistics

The focus of this study was on the comparison of the two treatments rather than estimating values to generalize to the population; therefore, it was imperative that the participants in the two treatments did not have differing characteristics which might be related to how individuals value the public goods of interest. After the PC exercise, study participants were asked a series of debriefing questions which focused on demographics, general attitudes, and respondents’ perceptions of the PC exercise. Tables II and III compare some of the demographic and attitude variables for the two groups.

Although the distributions of the age and education variables vary between the two treatments, the distributions are not judged to be statistically significant based on contingency table analysis ($\chi^2 = 32.94$, $p = 0.7022$ for age; $\chi^2 = 5.29$, $p = 0.8709$ for education). Likewise, the study participants in the two treatment groups had similar attitudes. Most of the participants rated themselves as liberal to moderate on social values (see Table III) and moderate to conservative on economic values. As some of the public goods were environmental amenities (specifically, open space and air quality), study participants were asked to rate how strongly they agree (or disagree) with the statement “I consider myself an environmentalist”. The majority

Table III. Attitudes of study participants.

<i>Variable</i>	<i>Treatment 1 (N = 53)</i>	<i>Treatment 2 (N = 50)</i>
<i>Social Values on a Political Scale</i>		
1 = Liberal	24%	18%
2	4%	6%
3	9%	4%
4 = Moderate	49%	45%
5	4%	4%
6	0%	2%
7 = Conservative	9%	20%
<i>Economic Values on a Political Scale</i>		
1 = Liberal	8%	12%
2	0%	2%
3	6%	6%
4 = Moderate	64%	50%
5	0%	8%
6	6%	6%
7 = Conservative	17%	16%
<i>I Consider Myself an Environmentalist</i>		
1 = Strongly disagree	2%	2%
2	15%	12%
3	15%	8%
4	13%	16%
5	9%	14%
6	26%	34%
7 = Strongly agree	19%	14%
<i>It was easy for me to choose between different public programs</i>		
1 = Strongly disagree	6%	4%
2	13%	10%
3	23%	12%
4	17%	4%
5	8%	34%
6	28%	26%
7 = Strongly agree	6%	10%

of the respondents in both treatments answered toward the strongly agree end of the scale. The one interesting difference between the two treatments was the response to the statement "It was easy for me to choose between different public programs". Treatment 1 participants were more likely to disagree with this statement which is encouraging given the public goods used in Treatment 1 were designed to be of

Table IV. Interpolated dollar values by treatment.

	Mean (Std. Dev.)	Median
<i>Treatment Group 1 (N=53)</i>		
AQ	\$251 (94)	\$300
OS	\$241 (96)	\$300
BIK	\$241 (100)	\$300
BUS	\$192 (120)	\$245
<i>Treatment Group 2 (N=50)</i>		
OS	\$227 (102)	\$300
TRL	\$188 (113)	\$220
RAP	\$112 (116)	\$48
BSH	\$ 69 (107)	\$18

similar value whereas those in Treatment 2 were designed to vary in value. The distributions of responses to the statement about ease of choosing between public goods were found to vary significantly for the two treatments based on contingency table analysis ($\chi^2 = 15.60, p = 0.0160$).

7. Testing for a Context Effect

The context did *not* significantly affect the estimated value of the open-space project in the sense that the measures of central tendency for the open-space project did not differ significantly between the two treatment groups. First we conducted the Lillifors test to see if the data were distributed normally. As the data were not found to be distributed normally, the nonparametric Mann-Whitney test was used to test for statistical differences in the distributions. The interpolated values were not found to be significantly different ($z = -0.7235, p = 0.4649$).⁸ The median value for OS in both treatments was \$300 so we removed the respondents in both treatments with interpolated values of \$300. Removal of these observations allowed us to compare the distributions of dollar amounts below \$300. Based on contingency table analysis ($\chi^2 = 11.943, p = 0.814$) we do not find evidence of the treatment influencing the distribution of responses.^{9,10} As mentioned in the previous section, the responses to the statement “It was easy for me to choose between the different PUBLIC goods” suggested participants in Treatment 2 responding toward the strongly agree end of the scale relative to Treatment 1 participants. We interpret this result as suggestive that it was more difficult for participants to make choices among goods that had similar value (i.e., Treatment 1). However, this difficulty did not seem to have a significant impact on the value estimates for the good common to the two treatments. The nature of the PC exercise with multiple binary choices in randomized order may avoid sensitivity of the value estimates to the choice set.

8. Testing for Scope

There is a clear difference in the estimated mean and median WTA values for the two bus programs with the estimated value of the more inclusive bus program (BUS) being significantly greater than that of BSH based on the Mann Whitney test ($z = -4.76$, $p = 0.0000$). Table IV shows the means and medians from the interpolated values. The strength of this result is encouraging evidence of the robustness of the value estimates derived from PC data. The percent of participants listing BUS as the *least* preferred good in choice set 1 (40%) is much less than the percent of participants listing BSH as the *least* preferred good relative to the other goods in choice set 2 (76%). Again, this shows that expansion of bus service for one weekend is less valuable to respondents than a permanent expansion of bus service.¹¹

9. Conclusion

The results of the experiments discussed in this paper suggest that the method of PC can provide robust measures of WTA.¹² This is an important first step to exploring alternatives to the standard CV approach for measuring WTA. The results of the two experiments discussed in this article suggest the method of PC is worthy of additional consideration as a viable approach to non-market valuation. The robustness of the value estimates across differing contexts and the sensitivity to scope are evidence of the construct validity of the PC measures. Ideally, we would like to test the relationship between the value estimates based on the method of PC and a criterion which is closely related to the relevant Hicksian surplus measure. Currently such tests are not practical. However, a next step might be to conduct experiments which allow us to compare the value estimates from the method of PC to those estimated from the standard CV approach. Loomis et al. (1996) conducted an experiment to compare WTP values estimated from a standard CV to WTA from the method of PC for a *market* good and found they were statistically different. However, the divergence between WTA and WTP was much smaller than has been found in most other hypothetical WTA-WTP comparisons and was closer to actual cash WTA-WTP comparisons. A natural extension would be to see how WTA values estimated from PC and CV data compare for a non-market good. We would expect the PC WTA values to be less than the standard CV WTA values as the PC approach avoids loss aversion. Experiments could be designed to estimate WTA via PC with loss aversion. Clearly more research needs to be conducted.

Notes

1. The PC exercise could also be conducted from chooser-reference point for losses, which would give a measure of WTP. It is also conceivable that PC could be used to measure WTA with loss aversion.
2. If costs are included as one of the attributes, the dollar value of the attributes can be inferred.

3. In an unpublished manuscript, Carson (1995) looks at the scope tests that have been conducted using the standard CV approach and concludes that contingent values are sensitive to scope if a properly designed and administered survey is conducted.
4. Both of these studies are highly controversial. The main criticism focuses on the quality of the survey instrument, specifically the information scenario, in both studies.
5. We first tested the data for normality. Given that the data were not normal, we used a nonparametric approach. The results of the both hypothesis tests do not change when means and medians are estimated from a parametric model.
6. The computer program was developed by George Peterson, Thomas Brown and Glenn Brink of the Rocky Mountain Experiment Station, USDA Forest Service, Fort Collins, CO.
7. The decision rule concerning how the public goods will be provided seems to influence the choices individuals make. Peterson et al. (1996) found that if the study participants were asked to be the sole decision maker about the provision of the public goods, the value estimates of the public goods were substantially higher than the situation where study participants were told the decision rule with respect to provision of the good was majority rule.
8. As mentioned earlier, the top dollar amount in the choice experiment was \$295. Several respondents preferred the open space project over \$295. Ideally, we should have included some higher dollar amounts. Since our pretests did not reveal this need prior to the final study, we assigned \$300 as the maximum for interpolating dollar values. The result is that our data set is skewed and the median value is \$300. If we were generalizing values, this would be a significant concern. However, given our goal of testing for a context effect, we do not feel that our conclusions would change if we used some other method of assigning maximum values.
9. Sixty-one values of \$300 were removed. Therefore, the contingency table analysis involved a total of 42 observations; 20 in treatment 1 and 22 in treatment 2. The Exact Tests (SPSS Manual 1996) procedure was used to account for the small sample size.
10. We are very grateful to one of the reviewers who suggested we conduct this additional test.
11. The ranking of goods as most preferred to least preferred was relative to the other goods in the choice set. Choice sets 1 and 2 contained different goods which would also influence the relative ranking.
12. WTA measured from the chooser reference point avoids loss aversion. We can't generalize to WTA with loss aversion.

References

- Bromley, D. W. (1995), 'Property Rights and Natural Resource Damage Assessments', *Ecological Economics* **14**, 129–135.
- Brown, T. C. and T. C. Daniel (1987), 'Context Effects in Perceived Environmental Quality Assessment: Scene Selection and Landscape Quality Ratings', *Journal of Environmental Psychology* **7**, 233–250.
- Carson, R. T. (1995), 'Contingent Valuation Surveys and Tests of Insensitivity to Scope', Unpublished Manuscript, Department of Economics, University of California, San Diego.
- Carson, R. T., N. E. Flores, K. M. Martin and J. L. Wright (1996), 'Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods', *Land Economics* **71**, 80–99.
- David, H. A. (1988), *The Method of Paired Comparisons*, Charles Griffin & Company, London.
- Desvousges, W.H., F. R. Johnson, R. W. Dunford, K. J. Boyle, S. P. Hudson and K. N. Wilson (1992), *Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability*, in *Contingent Valuation: A Critical Assessment*, Cambridge Economics, Inc.
- Fechner, G. T. (1860), *Elemente der Psychophysik*, Breitkopf and Härtel, Leipzig.
- Franciosi, R., P. Kajual, R. Michelitsch and R. Smith (1993), 'Experimental Tests of the Endowment Effect', unpublished manuscript.
- Hoehn, J. P. and J. B. Loomis (1993), 'Substitution Effects in the Valuations of Multiple Environmental Programs', *Journal of Environmental Economics and Management* **25**, 56–75.

- Kahneman, D. and J. Knetsch (1992), 'Valuing Public Goods: The Purchase of Moral Satisfaction', *Journal of Environmental Economics and Management* **22**, 57–70.
- Kahneman, D., J. L. Knetsch and R. H. Thaler (1990), 'Experimental Tests of the Endowment Effect and the Coase Theorem', *Journal of Political Economy* **98**, 1325–1347.
- Loomis, J., G. L. Peterson, T. C. Brown, P. Champ and B. Lucero (1996), 'Estimating WTA Using the Method of Paired Comparison and its Relationship to WTP Estimated using Dichotomous CVM', Proceedings 1996 W-133 Regional Research Meeting.
- Magat, W. A., W. K. Viscusi and J. Huber (1988), 'Paired Comparison and Contingent Valuation Approaches to Morbidity Risk Valuation', *Journal of Environmental Economics and Management* **15**, 395–411.
- National Oceanic and Atmospheric Administration, U.S. Department of Commerce (1993), Natural Resource Damage Assessments under the Oil Pollution Act of 1990. Federal Register 58, 4601–4614.
- Parducci, A. (1968), 'The Relativism of Absolute Judgments', *Scientific American* **219**, 84–90.
- Peterson, G. L. and T. C. Brown (1997), Economic Valuation by the Method of Paired Comparison with Emphasis on Tests of the Transitivity Axiom, Rocky Mountain Experiment Station Working Paper No. 97–2.
- Peterson, G. L., T. C. Brown, D. W. McCollum, P. A. Bell, A. A. Birjulin and A. Clarke (1996), 'Moral Responsibility Effects in Valuation of WTA for Public and Private Goods by the Method of Paired Comparison', in W. L. Adamowicz, P.C. Boxall, M.K. Luckert, W.E. Phillips, and W.A. White eds., *Forestry, Economics and the Environment*, CAB International, UK.
- Randall, A. and J. P. Hoehn (1992) 'Embedding Effects in Contingent Valuation: Implications for Natural Resource Damage Assessment', Staff paper 92-14, Department of Agricultural Economics, Michigan State University.
- Randall, A., J. P. Hoehn and G. S. Tolley (1981) 'The Structure of Contingent Markets: Some Results of a Recent Experiment', Paper presented at the annual meeting of the American Economic Association, Washington, DC.
- SPSS Exact Tests 7.0 for Windows (1996). Chicago, IL.
- Thurston, L. L. (1927) 'A Law of Comparative Judgment', *Psychological Review* **34**, 273–286.