Journal of Agricultural and Applied Economics, 34,1(April 2002):215–229

Applying a Method of Paired Comparisons to Measure Economic Values for Multiple Goods Sets

Randall S. Rosenberger, George L. Peterson, and John B. Loomis

ABSTRACT

A method of paired comparison is adapted for use in estimating economic measures of value. The method elicits multiple binary choices for paired items in a choice set. Probability distributions and economic values are estimated nonparametrically and parametrically. The method is applied in an experimental context with a choice set composed of four private goods and several sums of money. The sample's median value estimates for the goods are generally not different than the market prices for these goods. People who are in the market for a good value it higher than those not in the market for the good.

Key Words: *alternative gains, economic valuation, hypothetical market, paired comparisons, stated choice method.*

JEL codes: C51, C80, D12, Q26

Introduction

Psychologists, since the 19th Century, have developed and used the method of paired comparisons (PC) to elicit binary choices or judgments for paired items in a choice set. Respondents choose the item in the pair that has a greater magnitude on a given dimension, whether it is for physical properties such as weight or psychological properties such as preferences. Given z elements, the method presents them independently in pairs as $(z/2)(z-1)$ discrete binary choices.[1] The individual simply chooses the preferred element in each pair.[2] If there are no preference errors, and if preferences obey the axioms of utility

Randall S. Rosenberger is assistant professor, West Virginia University, Morgantown, and holds a co-appointment in the Regional Research Institute and Division of Resource Management. George L. Peterson is project leader, Rocky Mountain Research Station, Forest Service, US Department of Agriculture, Fort Collins, CO. John B. Loomis is professor, Department of Agricultural and Resource Economics, Colorado State University, Fort Collins.

This study was supported in part by funds provided by the Rocky Mountain Research Station, Forest Service, U.S. Department of Agriculture and the Social Sciences Institute, Natural Resource Conservation Service, U.S. Department of Agriculture. We gratefully acknowledge Andrea Clarke, Paul Bell, Andrej Birjulin, Dan McCollum, Patricia Champ, and Thomas Brown for their assistance in designing the paired comparison experiments. The manuscript benefited from insightful comments by anonymous reviewers for this journal. Any errors or omissions are the sole responsibility of the authors.

[1] When z is large, various methods can be used to reduce the number of choices presented to any individual (David; Green and Srinivasan).

[2] Whether to allow the individual an indifference option or to require a choice is debatable. We argue that forcing a choice in all cases maximizes discernment of difference while revealing indifference stochastically. Across individuals, or across repetitions of the choice for the same individual, the expected effect of indifference is an equal number of selections of each element in a pair. The requirement of a choice is similar to discrete choice contingent valuation format where the individual is allowed only two options, "yes" or "no".

theory (especially transitivity and comparability), the result will be a perfect rank ordering of the elements in the choice set. Since people can and do make choices from among subsets of the universal set of possible elements, the construction of a choice set should be sensitive to perceived permissible elements (Etzioni), choice-incentive compatability, and independent of irrelevant alternatives (Louviere).

Many arguments have been made concerning the need to be able to measure the economic value of goods and services, especially when these goods and services are not traded in established markets (Cummings, Brookshire, and Schultze; Mitchell and Carson; Peterson, Driver, and Gregory). Several valuation methods are in use, each with its own set of potential biases. Eliciting stated preferences through hypothetical markets is one way to estimate economic values. As in any science, it is prudent to continually test, expand and improve existing methods, and explore the application of new methods to current issues. One example is the recent extension of conjoint analysis (a stated choice method) from the marketing and transportation fields to environmental goods valuation (Gan and Luzar; Johnson and Desvousges; Roe, Boyle, and Teisl). Most stated choice methods are concerned with valuing the mix of attributes of a given good (Adamowicz *et al.*).

Conjoint analysis is the application of a multi-attribute technique that decomposes value sets of individual evaluations, or discrete choices, from a designed set of multi-attribute alternatives (Louviere). Conjoint analysis has been extensively used in marketing and transportation research and product development. In conjoint analysis there are essentially four different elicitation formats including ranking, rating, discrete choice, and graded-pair comparison applications (Louviere; Johnson and Desvousges). We use a discrete choice or paired comparisons method because it is the basic choice context. Choices elicited in a PC context can be used to derive ratings, rankings, and relative strengths of preferences for elements in a choice set.

We focus on using PC to value a mix of goods as opposed to the traditional application of conjoint analysis for valuing levels of attributes of a single good or program. For example, in a policy context we may be interested in ordering or measuring economic preferences for a variety of programs under resource constraints. Our application of PC begins with simple choice problems involving private goods. Broadened applications of PC to valuing public goods (Champ and Loomis; Peterson and Brown), to valuing varying levels of a good (Lockwood, 1998), or valuing changing levels of several attributes of a good, may blur a distinction between PC and conjoint analysis. The application of PC to valuing public goods would seem to be straightforward methodologically, as exemplified in the applications previously cited. However, many other observable and labile factors could affect people's preference expressions (Fischhoff; Schkade and Payne; Slovic). PC may be one method that could be used to investigate these factors and how they affect individuals' values for public goods (Lockwood, 1999).

PC has several potential advantages when compared with more traditional hypothetical market techniques for estimating economic values. These advantages can include: 1) merger of other disciplinary techniques (e.g., psychometrics) and knowledge with economics (Lockwood, 1999); 2) development of context-, decision-, and policy-relevant choice scenarios (Mitchell and Carson); 3) segmentation of affected heterogeneous parties (Swallow *et al.*); 4) reduction of framing bias in stated preference surveys (Mitchell and Carson); and 5) estimation of conservative willingness-to-accept compensation measures (Loomis *et al.*). One disadvantage of using PC models is the exacerbation of potential biases afflicting other hypothetical market valuation techniques, including information scenario development (Mitchell and Carson), use of heuristics in decision-making (Hogarth) and non-economic behavioral responses (Lockwood, 1999).

Even though our application of PC is with market goods, we have alluded to the potential application of the method to the valuation of non-market goods. We do not contend that the results of this experiment provide any evidence on the validity of PC to valuing nonmarket goods. However, if a method does not perform well when applied to goods with known markets, then we should have little confidence in its extension to contexts without markets. Our experiment regarding PC is a step back from other applications that have valued public goods (Champ and Loomis; Lockwood, 1998; Peterson and Brown). Our intent is to more fully develop the theory and econometric modeling of the data in the context of market goods. The rest of this paper discusses the following elements. The method of paired comparison is presented, including the type of economic value that is elicited, the experimental design, and different ways to structure the data collected. An experimental test of the method is discussed highlighting the applicability of the method to economic valuation. The data collected in the experiment is used to demonstrate how economic values can be estimated from the data. We will close with a few conclusions on the subject of the paper, and some recommendations for future research.

The Method of Paired Comparison and Economic Value

The relationship between preference ranking from a PC experiment and economic neoclassical utility theory has been presented elsewhere (Peterson and Brown). Suffice it to say, the intertwining among the roots of economics and psychological choice theories offers an opportunity to cast PC in terms of utility maximizing discrete choice theory, thus providing an economically consistent justification for the use of the well-developed psychometrics of PC to economic valuation applications.

PC gathers multiple trade-off observations for all elements in the choice set. If the goods in the choice set are well chosen, then PC may provide contexts and choice criteria for the revelation of theoretically consistent economic preferences. These contexts and choice criteria include an increase in the likelihood that the individual will think carefully about the characteristics, substitutability, trade-offs, and relative worth of the target good(s) in the course of making the required comparison. PC may also enable the identification of contexts in which the economic question is not perceived to be the appropriate question (Anderson; Sagoff).

The trade-off relationship between money (M) and a target good (X) can be defined using standard utility theory ($U(X,M)$). Three different reference points (buyer, seller, and chooser) and their corresponding valuation questions and subsequent value measures (WTP, WTA, and WTF)[3] can be identified (Kahneman, Knetsch and Thaler). First, if from the buyer reference point where the base or endowment case is $U(X^0,M^1)$ (does not possess the good (X^0)), the valuation question attempts to elicit maximum WTP for the gain in the good, or the question of where $U(X^0,M^1) = U(X^1,M^1 - WTP)$. Second, if from the seller reference point where the base case is $U(X^1,M^0)$ (possesses the good (X^1)), the valuation question attempts to elicit minimum WTA for the loss of the good, or the question of where $U(X^1,M^0) = U(X^0,M^0 + WTA)$. And third, if from the chooser reference point where the base case is $U(X^0,M^0)$ (does not possess the good), the valuation question attempts to elicit minimum WTF for not choosing the good, or the question of where $U(X^1,M^0) = U(X^0,M^1)$.[4] In either case, the measures (when one of the elements of a pair is money) are all monetary equivalents for the change in the

[3] WTP is willingness to pay from the buyer's reference point. WTA is willingness to accept compensation from the seller's reference point. WTF is willingness to accept from the chooser's reference point, or an individual's willingness to forego a gain in one element of a pair for the other element in the pair. Peterson and Brown call WTF "willingness to accept compensation from the chooser's reference point (WTA_c)". However, to avoid confusion with the traditional WTA measure from the seller's reference point, we use WTF to identify respondent choices between alternative gains.

[4] WTF reflects the opportunity cost of choosing one of the elements in a pair. In the case where one of the elements in a pair is money, then the opportunity cost for the good is the amount of money being foregone. The goal is then to elicit the minimum opportunity cost or WTF that equilibrates the amount of money being offered and the good. The exact relationship between WTF and WTA is an empirical question for which we have no data.

good (ΔX). Regardless of the reference point, all approaches are attempting to measure the equivalent change in income (ΔM) for X.

The chooser reference point question differs from the other reference point questions in that this question is concerned with two alternative future states and not a status quo state. In other words, the chooser reference point is when an individual chooses between alternative gains, selecting the one that maximizes her utility. For example, she may choose between $25 and a good, such as a cordless phone or a wildlife art print. By definition, we are attempting to measure the minimum amount of money she would be willing to accept to forego the gain in the relevant good offered. Similarly, the amount of money presented in each choice represents the opportunity cost of choosing the good—if she chooses the good, then her minWTF $\geq$ \$25; if the money is chosen, then minWTF $\leq$ \$25. Over the pairings of several sums of money with a good, we may be able to bound minWTF by the interval in which she switches from preferring the good over smaller sums of money to preferring larger sums of money over the good.

WTF is different from WTA in that no real or perceived loss from a good occurs in the chooser reference point scenario—the individual does not yet possess the good being offered in the experiment. This should avoid behavioral effects on WTA measures such as loss aversion (Kahneman and Tversky), resulting in economic measures that are more conservative than WTA measures. Kahneman, Knetsch and Thaler provide evidence that WTF is closer to WTP from the buyer reference point than WTA. Loomis *et al.* provide evidence that the ratio to WTP for the same good is smaller when WTF is estimated with PC than when WTA is estimated. Loomis *et al.* also show that this ratio of PC derived WTF to contingent valuation derived WTP is closer to ratios found in actual cash experiments. Thus, PC measures of economic value may be more conservative than measures of WTA from the traditional seller reference point, offering an alternative to the approach where WTP measures act as proxies for WTA measures.

Experimental Design

The paired comparison experiment was designed so that the binary choices or pairs of elements were randomly presented by means of a computer code. Each element in the experiment was identified by a short name. These names were presented in pairs that appeared side-by-side on the computer monitor. The computer program presented all possible pairings of the choice set elements. The participant simply had to choose the preferred element in each pair by pressing the right or left arrow key on the keyboard. Mistakes could be corrected by pressing the "backspace" key and changing the selection. The choice experiment was immediately followed by several debriefing questions including attitudinal and sociodemographic questions. The computer program recorded: 1) the participant's choice for each pair in an ordered matrix, which represents the participant's dominant preference order of the elements, 2) the time in seconds required for each choice, 3) the sequence number of each choice, 4) the pairs that resulted in circular triads or intransitivities (e.g., A>B>C>A), and 5) original choice switches, if any, for the two types of retrial choices—pairs that were identified to be part of a circular triad, and ten randomly selected pairs that were consistent with the participant's dominant preference order. The responses to these retrial pairs replaced the original choices in the ordered choice matrix.

Mitchell and Carson provide a typology of potential response effect biases in hypothetical valuation studies. Some of these biases can be reduced, if not eliminated, by the design of the PC computer experiment as defined above. Two key elements in any experimental design are randomization and replication (Thurstone). While PC cannot be made bias-free, several of the biases can be balanced out. These include starting point bias, importance bias (framing effect), position bias, and question order bias. The reduction of these biases was accomplished through a double randomized presen-

Table 1. Example of an Ordered Choice Matrix for Individual with Two Goods (A and B) and Four Sums of Money ($W, $X, $Y and $Z)

Elements	$W	A	$X	$Y	B	$Z	Row Sum
$W	—	1	1	1	1	1	5
A	0	—	1	1	1	1	4
$X	0	0	—	1	1	1	3
$Y	0	0	0	—	1	1	2
B	0	0	0	0	—	1	1
$Z	0	0	0	0	0	—	0
Column Sum	0	1	2	3	4	5	—

Note: The ordered choice matrix contains no circular triads. The sums of money are ranked according to dollar magnitude as $W < $X < $Y < $Z. Column Sum is a measure of the dominant preference order of the goods based on the number of times each element was chosen over all other elements in the choice set. Row Sum is a measure of the dominant preference order and is the inverse of the Column Sum.

tation of the pairs of elements according to sequence and placement of pairs on the monitor by the computer program. Question order and starting point bias are reduced or eliminated by randomizing the sequence in which the pairs are presented to each participant (Alberini). In addition, each participant had prior knowledge of the choice set in the experiment, reducing concern over pathway bias (Gregory *et al.*). Position bias is eliminated or reduced by randomizing the sequence of the pairs and the placement (right or left on the monitor) for each participant. Importance bias is potentially reduced through PC because of its ability to include several goods, providing not only issues of substitutability between goods, but also "hiding" the target good from the direct purview of the participant. However, with the inclusion of several goods, other biases, such as information bias, may be exacerbated by PC. Other forms of bias or uncertainty regarding the validity and reliability of value measures derived from experiments using hypothetical markets are issues that PC may not be an improvement on over other stated preference methods (Harris, Driver and McLaughlin).

The double randomization of the order and placement of each pair in the experiment allows each choice from every participant to be treated as independent from all other choices, including their own. If participants' preferences are well formed and known, then their preferences should be consistent and complete, whether presented randomly or not. Empirical evidence of this will be presented when the experiment is discussed in detail.

Structure of the Data

In PC, all elements of a pre-defined choice set are paired with all of the other elements, eliciting a response from each individual about their preferences for the elements in the choice set. Thus, the data in a PC experiment consists of n, $m \times m$ matrices where n is the number of people in the sample, and m is the number of elements in the choice set. Each observation, m_{ij}, is that person's binary response to the i^{th} and j^{th} element (with $i \neq j$). The m_{ij}'s can be coded as '0' being a "no" response (a preference for the j^{th} element over the i^{th} element), or as '1' being a "yes" response (a preference for the i^{th} element over the j^{th} element). The data can be partitioned in various ways for different analyses.

First, this choice matrix can be ordered for each participant by row or column sums, which provides a rank ordering of the elements in the choice set representing that participant's dominant preference order, or the number of times each element was chosen over all other elements in the choice set. Table 1 provides a visual description of a person's ordered choice matrix for a two goods by four sums of money experiment. This rank ordering can be used to explore the transitivity assumption of rationality in utility theory.

Second, all choice matrices, one for each participant, can be aggregated to provide the sample's aggregate ordered choice matrix where all participants receive a weight equal to one. Row and column sums of this matrix provide the sample's dominant preference order. This would be similar to a social rank ordering of the elements in the choice set.

Third, the choice matrix can be partitioned to isolate a participant's responses for a single element in the choice set. This is the extraction of the relevant choices from each row or column vector for the target good. If we further restrict the partitioning to only include responses between the target good and sums of money, and extract a row vector for each participant, then we would have a matrix (E) of the following form:

$$(1) \qquad E = [e_{ij} | s_{ik}] \quad \text{with} \begin{pmatrix} i = 1 \ldots n \\ j = 1 \ldots j \\ k = 1 \ldots k \end{pmatrix},$$

where n = sample size, j = sum of money, and k = attitude and sociodemographic variables.[5] The size of the matrix is $n \times (j + k)$, and the e_{ij} is the partitioned matrix of the binary coded responses as defined above. That is, all row vectors of 0's and 1's between the target element and the sums of money, one for each person, are stacked in a single matrix. This matrix is a collection of all participants' binary choices for the target good across all sums of money and attitude/sociodemographic variables (s_{ik}). This matrix provides us with the relevant choice information for estimating economic values for the target good. Table 2 provides a visual description for the two goods by four sums of money case with four individuals. Further data structuring is presented in the next section.

Estimating Economic Values from PC Data

Nonparametric Estimation

Economic values for each good in a PC experiment can be estimated parametrically or nonparametrically. For nonparametric measures, the relevant data matrix is the partitioned matrix that consists of all choices between a good and the sums of money as defined above (see Table 2 for an illustration). If each participant's preference rankings between a good and the sums of money are transitive and the elements are comparable and tradable, then the interval in which someone switches from preferring the good over a sum of money to preferring a sum of money over the good can be identified. For example, individual 2 in Table 2 prefers good B to \$W and \$X, but prefers \$Y and \$Z to good B, where \$W < \$X < \$Y < \$Z. The individual switches from preferring the good to preferring money in the interval \$X to \$Y. The economic value for the good lies in the interval bounded by the dollar amounts.

Estimating values for the sample can be accomplished in two ways. First, each participant's value can be aggregated and averaged as the sample's aggregate estimate. Second, each column of individual choices can be aggregated as the proportion of the sample preferring the good to a sum of money. Aggregate proportions can be mapped as the sample's empirical distribution or survival function (Kristrom). This empirical distribution is equivalent to a cumulative density function. The sample's mean and median value for the good are directly derivable; the former is simply the integration under the empirical distribution function, and the latter is the value cor-

Table 2. Example of a Goods Matrix Consisting of Four Individuals' Binary Choices Between Good B and Four Sums of Money (\$W, \$X, \$Y and \$Z)

Individual	GOOD B			
	\$W	\$X	\$Y	\$Z
1	0	0	0	1
2	0	0	1	1
3	1	1	1	1
4	0	1	0	1

Note: The sums of money are ranked according to dollar magnitude as \$W < \$X < \$Y < \$Z. Individual four illustrates a circular triad.

[5] The attitude/sociodemographic variables, along with other information collected on each individual, can be appended to each row vector.

responding to the 50 percent survival level (Kristrom).

These nonparametric estimates, however, are point estimates of the value of the good. There may be circumstances when it is important to be able to estimate the distribution function and the corresponding mean and median values parametrically. Parametric models can also capture information on covariate effects of measured observable (e.g., income) and labile (e.g., attitudes) differences between individuals or classes of individuals. Parametric models can then be used to predict estimates, transfer estimates to other sites (benefit transfer), or generalize estimates for other populations.

Parametric Estimation

Each choice in a PC experiment is a discrete binary choice. Therefore, standard limited dependent variable analysis can be used to estimate a cumulative density function and subsequent mean and median values. Two standard approaches include Hanemann's utility difference approach and Cameron's compensation function. McConnell has shown that these two approaches are dual with linear specifications of the random utility model and constant marginal utility of income. Hanemann's approach is adopted as a matter of computational convenience. In this approach, participants are believed to be using a random utility difference approach when deciding which element to choose out of a pair of elements presented to them. Similar to discrete choice contingent valuation applications, the money element in each pair represents the "offer bid" (*$BID*). If the utility difference is logistically distributed,[6] a logit model of the probability of choosing a sum of money over a good is related to the sum of money (*$BID*) and attitude/sociodemographic variables (S_k) as in the following equation:

$$(2) \qquad R_i = \beta_0 + \beta_1(\$BID) + \beta_k(S_k)$$

where R_i is the binary response variable for pariticipant *i*. The probability of a "no" response and a "yes" response are, respectively:

$$(3) \qquad \mathrm{Pi}^{n}_{\$BID} = g(\$BID, S_k), \quad \text{and}$$

$$(4) \qquad \mathrm{Pi}^{y}_{\$BID} = 1 - g(\$BID, S_k),$$

where $g(\$BID)$ is the cumulative density function. The mean value is the area under this function as defined by:

$$(5) \qquad C^{+} = \int_0^{\infty} [1 - g(\$BID, S_k)]\, d(\$BID, S_k).^{7}$$

Several econometric regression techniques can be performed on the different structures of the PC data. The data matrix consisting of each participant's choices between a good and the sums of money can be restructured by stacking each person's responses in a column vector with all other participants' responses. For example, in Table 2, a column vector of all data points in the matrix can be constructed. A logit regression can then be estimated on this data structure.

The method above treats the multiple responses for each person as independent. There are, however, several forms of independence. The form we are specifying here is in the context of the experimental design. It is doubtful that any person took cues from prior choices in the experiment because of the volume of choices made in the brief amount of time required to make the choices. There are no systematic effects of question order in the aggregate due to the randomization. However, the multiple responses from each person may not be independent in the context of that person's value of the good.

It is reasonable to assume independence of

[6] Empirical evidence suggests that values obtained in hypothetical valuation experiments have a measurement error that is log-normal distributed (Rowe, Schulze and Breffle). We also tested a normal distribution for our data, but found the log-normal distribution of error had a better fit on the data than the normal distribution.

[7] In certain cases a truncated mean estimate may be required. This can be estimated by truncating the integration of the density function at the maximum bid offer in the experiment.

responses to meet the goal of this analysis. We have already implicitly assumed independence of responses in estimating the empirical distribution from the aggregate proportions for the sample. The goal of the regression analysis is to predict this distribution parametrically. A binary logit treatment of the PC data results in $n \times j$ data points or observations, where n is the sample size and j is the number of sums of money in the experiment. This treatment of the data does not make any assumptions concerning the triviality of information on choices above or below the participant's switching interval (such as is the case in a multiple-bounded treatment), including multiple intervals for those exhibiting inconsistent preference orders (e.g., individual 4 in Table 2). Degrees of freedom are also increased.

Other regression techniques are appropriate if the data are restructured or the dataset is reduced in size. For example, the matrix previously described can be treated as panel data (multiple responses from each person), and a relevant technique employed, such as random effects or fixed effects models (Rosenberger and Loomis). A multiple-bounded regression technique for double-bounded data can be used when each person's responses remain as a row vector in the aggregate response matrix. This technique searches the data and uses only the information on the relative switching interval for each person (Loomis *et al.*). Inconsistencies in individual preference orderings may pose problems for this method. Both of the methods, random effects and multiple-bounded logit, were tested. However, these methods did not estimate the sample's aggregated proportions distribution function as well as the logit analysis.[8]

Application

Our experiment consisted of four private goods and 12 sums of money in the choice set. The four private goods (and market prices) were a signed wildlife art print ($35), a cordless phone ($80), a certificate for dinner and beverage for two at a local restaurant ($30), and two tickets to a local college football game (seats on the 20 yard-line) ($30). None of the participants was told the market price for these goods at any time during the experiment. The 12 sums of money were $4, $8, $12, $18, $24, $30, $38, $48, $72, $120, $195, and $295. The sums of money were not compared to each other, so each individual made at least 54 binary choices, for at least 5,562 binary choices for the sample.[9]

Clarification of instructions and the setting of the sums of money were completed through several pretests of university staff. The four private goods were described on a "product sheet" that was given to each individual and could be referred to at any point in the experiment. In addition, the wildlife art print and the cordless phone were displayed in the room, and the dinner and football tickets were mounted on poster board in the room. The actual goods and poster board were on display during the experiment. The target good for this experiment was the art print. The signed art print was of a gray wolf in natural surroundings.

The sample consisted of college clerical and administrative staff in academic and nonacademic units at a land grant university. They were compensated $20 each for attending one of the 45-minute sessions. These sessions were conducted before work, at lunchtime, and after work. The total sample size for this experiment was 103 people distributed across the 14 different sessions. The constraint of number of computers available determined the size of each session.

[8] The rule used to determine "best fit" consisted of minimizing mean square error of the estimated proportions from the empirical aggregate proportions. This rule was chosen because it identifies which parametric model fits the empirical distribution of the data. Specific results are available from the authors upon request.

[9] More than 5,562 binary choices in the experiment were obtained since any pairs that were determined to be involved in a circular triad (intransitivity) and ten randomly selected consistent pairs were retried at the end of the original iteration of pairs in the experiment. The exact number of retrial pairs depends on the consistency of the individual's original choices. Circular triads, or cases where A>B>C>A, were identified by the individual's dominant preference ordering as determined by her overall choices in the original pairings of the choice set elements.

A principal investigator, following a written script, led all sessions. All sessions used the experimental format. The investigator led the participants through the experiment and provided written instructions on using the computers. The computer program presented the PC experiment first, followed by debriefing questions, which included attitudinal and sociodemographic questions. The exact wording of the introduction to the PC exercise was:

> "When the choice appears on the screen, please choose the one that you would like to receive if it were to be actually offered to you. Consider each choice independently, as if it were the only choice you had to make. While these choices are hypothetical and you will not actually receive either of the goods, make your choices as if you would actually receive one of the two goods."

All participants that began in a session finished the session even though they were told in writing that they were free to leave the session at any time, with compensation. The participants were careful in following the instructions and did not discuss their choices with others during the session. Comments after the sessions suggested that the participants were stimulated by the experience.

Results

Consistency of Preferences

The consistency of responses can be measured by the number of circular triads in the individual's dominant preference order of the choice set elements. A circular triad is when choices imply an intransitive preference rank ordering (A > B > C > A), such as individual 4 in Table 2. Circular triads may be caused by systematic intransitive preferences, random choice in cases too close to call, incompetence of the respondent, or simple mistakes. Intransitivities are systematic and repeatable circular triads, whereas inconsistencies are non-repeatable (David).

Utility theory postulates that rational individuals will have well-known and transitive preferences. The results of the experiment suggest that people's preferences are consistent, transitive, and reliable. In the total set of 5,562 initial choices, 186 (3.3%) of the choices were involved in circular triads, and 5,376 (96.7%) were consistent with each individual's dominant preference order. Across the entire sample, 52.7% of the inconsistent pairs were switched on retest, with only 6.9% of the ten randomly selected consistent choices switched on retest. These results imply that most circular triads are the result of mistakes or pairs whose value is too close to call, and do not imply intransitivity of preferences. In the case of pairs with equal value, each element in a pair would have a 50% chance of being selected.

Peterson and Brown investigated the issue of value difference between pairs on decision time and reliability of preferences in a mixed, private goods-public goods PC experiment. The results of their investigation show that the smaller the value differences between the elements in a pair, the higher the probability of indifference leading to inconsistencies. They also found that decision time increased when the value of the elements were close, implying harder choices and increased chances of inconsistency. The discovered stability of respondent's choices on re-tested consistent choices implies that the individuals' preferences are reliable.

Economic Value Estimation

Economic value estimates are derived from both the empirical distributions and the logit estimated distributions for each good in the experiment. The estimated distributions are based on binary logit regressions as defined above. All observations, including people who exhibited intransitivities in the form of multiple switching intervals, are included in the datasets. There are 24 people exhibiting intransitivities in the wildlife art print dataset, 15 individuals in the dinner tickets dataset, 14 individuals in the football tickets dataset, and 11 individuals in the cordless phone dataset.

We also asked each respondent if he/she was in the market for the goods being offered. Specifically, we asked whether each individual

Table 3. Estimated Binary Logit Models for Each Good[a]

	Variable Coefficients					Summary Statistics	
Good	Constant	*$BID*	*MKT*	*AGE*	*INC*	χ^2	Log-likelihood
Art Print	−2.520*	1.546*	−1.636*	−0.057*	0.002	636	−537
	(−6.29)[b]	(17.16)	(−9.81)	(−6.71)	(0.68)		
Dinner Tickets	−3.024*	0.077*	−1.126*	0.026*	0.005	736	−460
	(−7.66)	(14.05)	(−6.47)	(3.03)	(1.50)		
Football Tickets	−1.458*	0.061*	−1.067*	0.013	−0.005	580	−512
	(−4.06)	(12.86)	(−5.89)	(1.54)	(−1.71)		
Cordless Phone	−1.336*	0.042*	−0.883*	−0.014	−0.011*	814	−412
	(−3.37)	(15.14)	(−4.78)	(−1.42)	(−2.82)		

[a] The generic equation estimated is Prob(y) = $\beta_0 + \beta_1\$BID + \beta_2 MKT + \beta_3 AGE + \beta_4 INC$. The dependent variable is the proportion of the sample rejecting the good offered. The sample size is 103 individuals with 12 responses each, for a grand sample size of 1,236 binary responses.

[b] Asymptotic t-statistics reported in parenthesis.

* Significant at the 0.01 critical level.

was in the market for or had considered buying wildlife art prints or cordless phones, respectively. We also asked if each individual was in the market for or frequently bought football tickets or dinner at the local restaurant specified, respectively. Over 67 percent of the sample stated they were in the market for a cordless phone. About 40 percent of the sample stated they were in the market for the wildlife art print or the dinner tickets. Twenty-seven percent of the sample stated they were in the market for the football tickets.

A binary logit regression was performed on each data set, and included a constant term, and four independent variables—bid offered (*$BID*), a dummy variable for 1 = in the market for the relevant good and 0 = not in the market for the relevant good (*MKT*), age of the respondent in years (*AGE*), and gross household income in $1,000 (*INC*). The results from the regressions are presented in Table 3.

The constant term, *$BID* and *MKT* are significant in all four models. *AGE* and *INC* are significant in some of the models. The sign of the estimated coefficient on *$BID* shows that as the sum of money offered increases, the probability of rejecting the good in favor of the money increases. The sign on the *MKT* variable shows that people who are in the market for the good have a lower probability of rejecting the good in favor of the money. In other words, people who are in the market prefer the good and thus have a higher opportunity cost of not choosing it than people who are not in the market. Information on "market" involvement of survey participants in stated preference economic surveys would prove beneficial in understanding the strength of people's preferences for goods or services. Without this information, stated preference surveys treat all individuals the same, ignoring the importance of market segmentation and implied strength of budget constraints.

The sign of the coefficient on *AGE* shows that older people have a higher WTF for the art print and the cordless phone than younger people, whereas younger people have a higher WTF for the dinner and football tickets than older people. The estimated coefficients on *INC* show similar results. People with higher household incomes have a higher WTF for the football tickets and cordless phone than lower income households, whereas lower income households have a higher WTF for the art print and dinner tickets than higher income households.

Table 4 reports the aggregate proportions of the sample based on the observed levels from the raw data and the predicted levels based on

Table 4. Observed and Predicted Aggregate Proportions of the Sample Choosing the Good Over the Sums of Money

	Art Print		Dinner Tickets		Football Tickets		Cordless Phone	
Sum of Money	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.
$4	0.88	0.96	0.92	0.98	0.84	0.95	0.99	1.00
$8	0.85	0.89	0.88	0.92	0.82	0.84	0.97	0.99
$12	0.82	0.81	0.85	0.82	0.75	0.72	0.95	0.97
$18	0.71	0.70	0.69	0.64	0.56	0.56	0.94	0.93
$24	0.64	0.60	0.52	0.48	0.47	0.43	0.88	0.86
$30	0.54	0.52	0.41	0.35	0.30	0.33	0.85	0.78
$38	0.41	0.42	0.27	0.23	0.25	0.24	0.76	0.67
$48	0.35	0.34	0.11	0.15	0.11	0.17	0.60	0.53
$72	0.22	0.22	0.03	0.06	0.07	0.09	0.35	0.30
$120	0.12	0.11	0.00	0.02	0.02	0.04	0.12	0.11
$195	0.09	0.06	0.00	0.01	0.02	0.02	0.05	0.04
$295	0.05	0.03	0.00	0.00	0.01	0.01	0.02	0.01
% Correctly predicted 0's		79		80		68		93
% Correctly predicted 1's		81		90		89		80

the estimated logit functions. The aggregate proportions reported are the proportions of the sample choosing the good over the sum of money. Figures 1 and 2 graph the nonparametric and parametric (logit-estimated) distributions, respectively, for all of the goods in the choice set. The parametric distributions nearly perfectly map the nonparametric distributions (Figure 3). Adjusted-R^2 goodness-of-fit statistics between the nonparametric and parametric aggregate proportions are in excess of 0.98 for all four models. The probability of each model correctly predicting "0" responses ranged from 0.68 for the football tickets to 0.93 for the cordless phone (Table 4). The probability of each model correctly predicting "1" responses ranged from 0.80 for the cordless phone to 0.90 for the dinner tickets (Table 4).

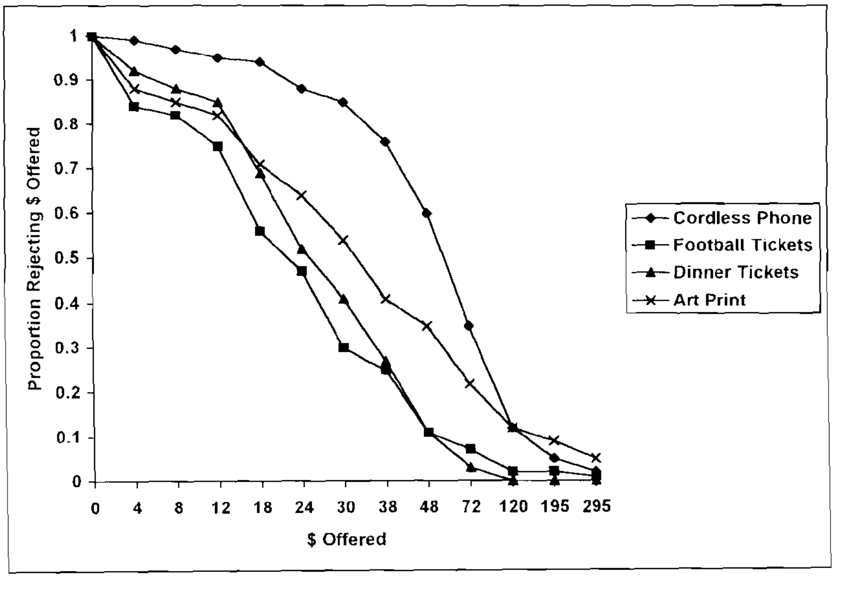

Figure 1. Nonparametric distribution functions for all goods in the choice set

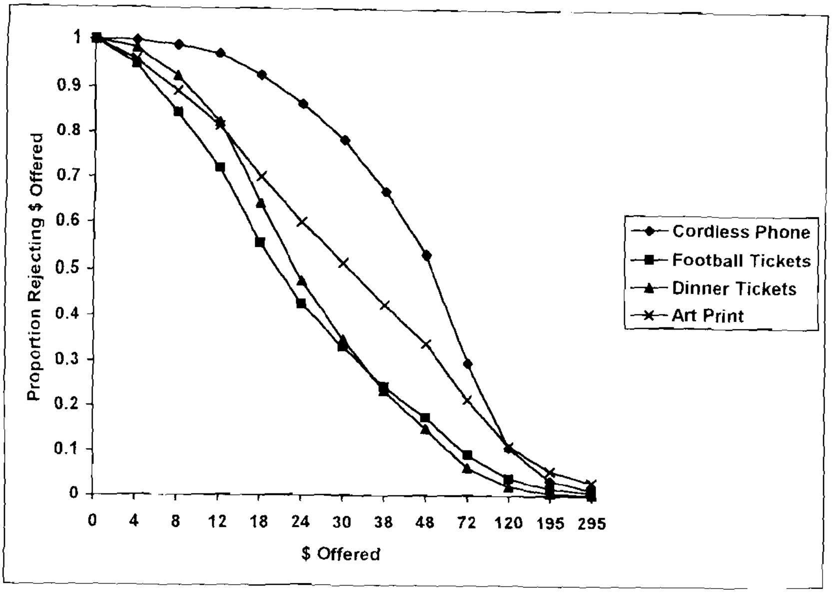

Figure 2. Logit estimated distribution functions for all goods in the choice set

Both the mean and median WTF values for the goods were estimated from both models for each good. While the mean value may better represent the average WTF of the good in question for economic evaluation purposes, the median value is probably more representative of a market value than mean WTF. The median measures the amount of money that would make at least half of the sample switch from choosing the good to accepting the money. The median value is also less sensitive to a few individuals that have extreme values for the good. We will only discuss the median measures here. Table 5 reports both the median and mean measures of WTF.

For the art print, the median WTF was es-

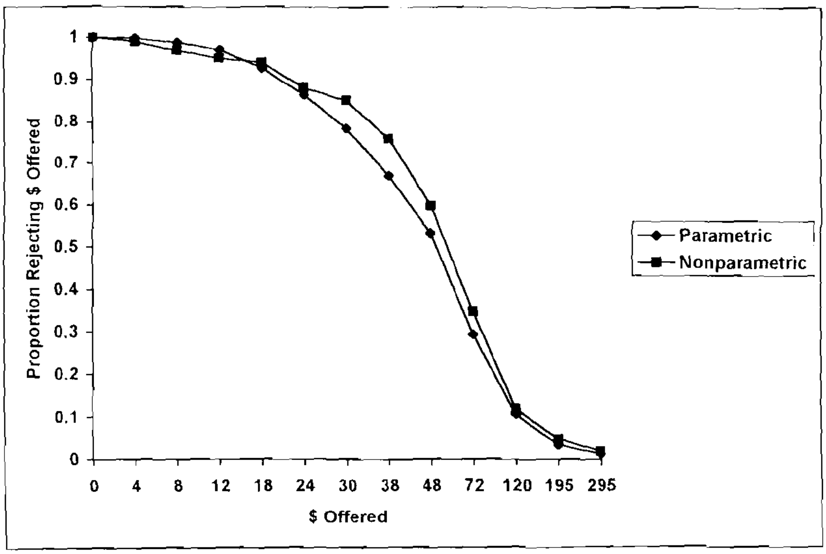

Figure 3. Aggregate proportions distribution and logit estimated distribution functions for the cordless phone

Table 5. Mean and Median WTF Estimates for Each Good in the Experiment[a]

Measures	Art Print	Dinner Tickets	Football Tickets	Cordless Phone
Market Price	\$35	\$30	\$30	\$80
Median WTF				
Nonparametric Model	\$32	\$26	\$23	\$58
Parametric Model				
Full Sample	\$32	\$28	\$26	\$65
In the Market	\$60	\$37	\$37	\$78
Not in the Market	\$21	\$23	\$19	\$57
Mean WTF				
Nonparametric Model	\$55	\$30	\$37	\$75
Parametric Model				
Full Sample	\$72	\$30	\$29	\$66
In the Market	\$135	\$38	\$38	\$78
Not in the Market	\$47	\$25	\$24	\$59

[a] The mean and median WTF values are estimated from the parametric logit models using Hanemann's formulas.

timated to be \$32 by both the nonparametric model and the parametric model. The market price for the art print was \$35. Individuals that stated they were in the market for art prints would be WTF \$60 for the art print, whereas those that stated they were not in the market would be WTF up to \$21 for the art print. The dinner tickets had a market price of \$30. The nonparametric model predicted \$26 and the parametric model predicted \$23 for the dinner tickets. Those individuals that stated they were in the market for dinner tickets would be WTF up to \$37, whereas those that stated they were not in the market would be WTF only up to \$23 for the dinner tickets. The football tickets were estimated to be worth \$23 based on the nonparametric model and \$26 based on the parametric model. The market price for the football tickets was \$30. The 'in the market' segment would be WTF up to \$37, while the 'not in the market' segment would be WTF only \$19 for the football tickets. The cordless phone had the highest WTF value of all the goods in this experiment. Its median WTF value was estimated to be \$58 from the nonparametric model and \$65 from the parametric model. Its market price was \$80. For those individuals stating they were in the market, they would be WTF up to \$78 for the phone. For those individuals stating they were not in the market for a phone would be WTF up to \$57 for the phone.

Discussion and Conclusions

We presented the application of a method of paired comparison (PC) to the estimation of economic values for private goods with known market values. The predicted median values were very close to the market price for each good. The economic value measure is from the chooser reference point. That is, the measure is the opportunity cost of not choosing an alternative gain. Each participant in the experiment provided a choice between all possible pairs of elements in a choice set that included sums of money. The multiple responses from each person allow identification of each person's relative preference order among the elements in the choice set, and their nonparametric survival function. Aggregating across the sample enables the identification of the sample's nonparametric cumulative distribution function based on aggregate proportions. Parametric modeling of the sample's cumulative distribution function is also possible given different structuring of the data.

We also suggested other advantages to using PC over conventional hypothetical market valuation methods. These advantages are in part due to the computer-aided application of the method. PC provides an efficient manner to collect multiple binary responses from each individual in a sample. There was no evidence of respondent fatigue in the experiment with at least 54 binary choices per individual. The data collected also provides evidence in support of the assumption of transitivity in utility theory, and that people's responses are reliable and consistent. The randomized presentation of pairings of elements reduces, or in the best possible case, eliminates the effect of bias in the results, especially question order bias, starting point bias, position bias, and importance bias (framing effect). However, the size of the choice set possible in a PC experiment exacerbates other sources of bias, such as information bias. For example, an extension of the choice set to include unfamiliar or complex goods such as environmental goods or social services requires additional information regarding these elements.[10]

We conclude that PC is a useful way to explore individual economic choice behavior and shows promise for application to hypothetical market estimation of economic values, including conservative measures of WTA. The richness of information obtained through this method and the ease of application encourage further research of its potential. This research can be on three different dimensions: 1) expanding on the existing applications of the method for monetary value estimation, including valuing non-market goods and services, 2) testing the validity and sensitivity of the method, and 3) applications of the method to elicit and estimate values in non-monetary terms.

References

Adamowicz, W., P. Boxall, M. Williams, and J. Louviere. "Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation." *American Journal of Agricultural Economics* 80(1998): 64–75.

Alberini, A. "Efficiency vs. Bias of Willingness-To-Pay Estimates: Bivariate and Interval-Data Models." *Journal of Environmental Economics and Management* 29(1995):169–180.

Anderson, E. *Value in Ethics and Economics.* MA: Harvard University Press, 1993.

Cameron, T.A. "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(1988):355–379.

Champ, P.A., and J.B. Loomis. "WTA Estimates Using the Method Of Paired Comparison: Tests of Robustness." *Environmental and Resource Economics* 12(1998):375–386.

Cummings, R.G., D.S. Brookshire, and W.D. Schulze (eds.). *Valuing Public Goods: An Assessment of the Contingent Valuation Method.* NJ: Rowman & Allanheld, 1986.

David, H.A. *The Method of Paired Comparisons.* London: Charles Griffin and Co., 1988.

Etzioni, A. *The Moral Dimension: Toward a New Economics.* NY: The Free Press, 1988.

Fischhoff, B. "Value Elicitation: Is There Anything in There?" *American Psychologist* 46(1991): 835–847.

Gan, C., and E.J. Luzar. "A Conjoint Analysis of Waterfowl Hunting in Louisiana." *Journal of Agricultural and Applied Economics* 25(1993): 36–45.

Green, P.E., and V. Srinivasan. "Conjoint Analysis in Consumer Research: Issues and Outlook." *Journal of Consumer Research* 5(1978):103–123.

Gregory, R., J. Flynn, S.M. Johnson, T.A. Satterfield, P. Slovic, and R. Wagner. "Decision-Pathway Surveys: A Tool for Resource Managers." *Land Economics* 73(1997):240–254.

Hanemann, W.M. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." *American Journal of Agricultural Economics* 66(1984):332–341.

Harris, C.C., B.L. Driver, and W.J. McLaughlin. "Improving the Contingent Valuation Method: A Psychological Perspective." *Journal of Environmental Economics and Management* 17(1989):213–229.

Hogarth, R. *Judgement and Choice,* 2nd Edition. NY: John Wiley and Sons, 1987.

Johnson, F.R., and W.H. Desvousges. "Estimating Stated Preferences with Rated-Pair Data: Environmental, Health, and Employment Effects of Energy Programs." *Journal of Environmental Economics and Management* 34(1997):79–99.

Kahneman, D., and A. Tversky. "Prospect Theory:

[10] See Peterson and Brown for an application in which public goods are included in the choice set.

An Analysis of Decision under Risk." *Econometrica* 47(1979):263–291.

Kahneman, D., J.L. Knetsch, and R. Thaler. "The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives* 5(1991):193–206.

Kriström, B. "A Non-Parametric Approach to the Estimation of Welfare Measures in Discrete Response Valuation Studies." *Land Economics* 66(1990):135–139.

Lockwood, M. "Integrated Value Assessment Using Paired Comparisons." *Ecological Economics* 25(1998):73–87.

———. "Humans Valuing Nature: Synthesizing Insights from Philosophy, Psychology, and Economics." *Environmental Values* 8(1999):381–401.

Loomis, J.B., G.L. Peterson, T.C. Brown, P.A. Champ, and B. Lucero. "Paired Comparison Estimates of Willingness to Accept versus Contingent Valuation Estimates of Willingness to Pay." *Journal of Economic Behavior and Organization* 35(1998):501–16.

Louviere, J.J. "Conjoint Analysis Modeling of Stated Preference: A Review of Theory, Methods, Recent Developments, and External Validity." *Journal of Transportation Economics and Policy* 22(1988):93–119.

McConnell, K.E. "Models for Referendum Data: The Structure of Discrete Choice Models for Contingent Valuation." *Journal of Environmental Economic and Management* 18(1990):19–34.

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

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

Peterson, G.L., B.L. Driver, and R. Gregory (eds.). *Amenity Resource Valuation: Integrating Economics With Other Disciplines.* PA: Venture Publishing, 1988.

Roe, B., K.J. Boyle, and M.F. Teisl. "Using Conjoint Analysis to Derive Estimates of Compensating Variation." *Journal of Environmental Economics and Management* 31(1996):145–159.

Rosenberger, R.S., and J.B. Loomis. "Panel Stratification in Meta-Analysis of Economic Studies: An Investigation of Its Effects in the Recreation Valuation Literature." *Journal of Agricultural and Applied Economics* 32(2000):459–470.

Rowe, R.D., W.D. Schulze, and W.S. Breffle. "A Test for Payment Card Bias." *Journal of Environmental Economics and Management* 31(1996):178–185.

Sagoff, M. "Aggregation and Deliberation in Valuing Environmental Public Goods: A Look Beyond Contingent Pricing." *Ecological Economics* 24(1998):213–230.

Schkade, D.A., and J.W. Payne. "How People Respond to Contingent Valuation Questions: A Verbal Protocol Analysis of Willingness to Pay For an Environmental Regulation." *Journal of Environmental Economics and Management* 26(1994):88–109.

Slovic, P. "The Construction of Preference." *American Psychologist* 50(1995):364–371.

Swallow, S.K., T. Weaver, J.J. Opaluch, and T.S. Michelman. "Heterogeneous Preferences and Aggregation in Environmental Policy Analysis: A Landfill Siting Case." *American Journal of Agricultural Economics* 76(1994):431–443.

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