Contingent Valuation and the Method of Paired Comparisons

David C. Kingsley
August 2009

Abstract

It has long been recognized that the total value of natural resources includes both direct and passive use values. The only method available to researchers interested in the total value associated with nonmarket goods is the contingent valuation method (CVM). Many government regulations require that a benefit cost analysis be performed when changes are proposed to environmental resources. This requires the frequent use of the CVM in public policy debates over policy alternatives. The correct measure of value, Willingness to Accept or Willingness to Pay, depends on the implied property rights associated with the change in the environmental resource. However, using the CVM the WTA has been difficult to measure reliably and researchers have focused on using WTP as a proxy for this measure despite possible bias. The paired comparison method (PCM) has been shown to reliably measure WTA. Further, it has been suggested that consumers in unfamiliar contingent markets require experience and familiarity with the environmental resource in order to express meaningful preferences. The PCM offers respondents such experience and has been shown to be consistent with preference learning models. As a result the PCM offers several advantages over traditional CV methods. This paper builds a novel econometric model which takes advantage of the data available within paired comparison data to offer a model with great flexibility that is familiar to random utility practitioners. Results suggest that PC data can be adapted to these methods and offers researchers an alternative CV method for estimating individual WTA.

David C. Kingsley is an Assistant Professor of Economics at Westfield State College, Westfield MA 01085. Email: dkingsley@wsc.ma.edu
Keywords: Nonmarket Valuation, Cost-Benefit Analysis, Choice Experiments
Acknowledgment: This paper benefitted from discussions with Tom Brown, Patricia Champ, Nick Flores and Donald Waldman. All errors remain my own
Working draft - comments welcome - do not cite without consent.
Introduction

Policymakers must often choose between several public policy alternatives. For instance, possible alternatives may include air pollution mitigation, environmental hazard cleanup or the creation of public space. In order to properly judge the efficacy of a public policy proposing changes in environmental resources a full benefit cost analysis must be employed. This implies that the total value of the change must be measured as well as the total cost. Total value is recognized to include both direct and passive use values. The recognition of the possibility of substantial passive use values within total value can be traced back to Kutrilla’s work in the 1960’s.

The Contingent Valuation method is the only nonmarket valuation technique able to measure the passive use component of total value. Without the CVM passive use values could not be directly measured and would only be included in official benefit cost decision-making through the political process or through expert consideration. Ignoring such values would almost certainly tilt policy away from preservation and towards a reduction in environmental goods as benefits would be systematically underestimated.

Estimating the benefits associated with changes in a natural resource is done using one of two common welfare measures, Willingness to Accept (WTA) or Willingness to Pay (WTP). The proper measure may be defined by WTA or WTP depending on the property rights. For example, consider these measures when the proposed change is an environmental good or improvement. When it is believed that individuals have the right to the status quo, no change, level of utility then the WTP is the correct measure of benefits associated with an environmental improvement. This implies that consumers are not entitled to the environmental good and would be required to pay, increased taxes, in order to obtain it. However, if individuals have the right to the change, or improved environmental good, the correct benefit measure is WTA. In other words individuals should be compensated for the loss of the environmental good as they are entitled to it. However, estimates of individual WTA have been difficult to reliably measure [24] (Brookshire, Randall Stoll 1980).

Despite the frequent use of the CVM to measure the economic value of nonmarket goods its use remains controversial. There are several well established biases related to the method notably hypothetical bias. Properly specified CVM studies using
best practice often mitigate these biases but questions remain. With the continuing debate concerning CVM economists have sought out new methodologies to measure total value of changes in natural resources. The paired comparison method offers a promising stated preference alternative with several advantages over traditional CV methods.

Paired comparison data involves a sequence of choices between pairs of items and is unique among contingent valuation applications for several reasons. Pairs may include locally relevant public goods, familiar private goods or dollar amounts. It significantly improves upon the context within which the choice is made. The tradeoffs between public policy alternatives is made explicit and should act as substitute reminders as well as providing policy makers with important rankings of the publics’ priorities. The choices between private goods and money are familiar providing consumers with a benchmark with which to make the more difficult choices between two public goods or a public good and a monetary amount. This choice format is easy as it replicates the closed ended format preferred by economists and it avoids asking for explicit dollar values on public goods. Importantly it allows researchers to retest, within individual, choices deemed to be inconsistent. If, as suggested by previous PC research, respondents become better able to discriminate between choices as they progress through a PC experiment then these retested choices may more accurately reflect underlying preferences. It is well understood that becoming in touch with preferences, particularly for public or unfamiliar goods, is a difficult process. Respondents require experience and familiarity in order to express well founded preferences (Plott, 1996). Offering consumers the ability to make such tradeoffs with a PC experiment may force consumers to focus on these underlying preferences.

The primary shortcoming of the PC data described in this paper is plausibility of payment mechanism and lack of description for the goods. As described it is a choice between two objects including monetary amounts, there is no mention of payment. Additionally the goods are only described briefly rather than exhaustively in similar CV designs. Both major weaknesses are easily addressed in future work. This paper lays the groundwork for future endeavors into the paired comparison method of contingent valuation.

This paper will develop novel econometric models in order to analyze PC data within a traditional random utility framework familiar to most economists. First, a
heteroscedastic probit model allows for individual specific error terms while allowing for
another level of heteroscedasticity over choice sequence. The model is able to control for
the influence of covariates across items allowing policymakers to measure how different
segments of the public value each alternative. This paper explicitly takes advantage of the
full range of data available within these experiments including the choices made between
items which provide data on relative values and the retested choices which may reflect
refined preferences.

Results suggest that the method of paired comparisons is a promising contingent
valuation methodology when a WTA is sought and multiple public policy alternatives are
being considered. Future work must rigorously test the validity and reliability of the PC
method similar to the tests that other CV methods have undergone.

**Contingent Valuation**

Benefit cost analysis is used to judge the efficacy of public policy without
controversy in a great deal of situations. Often there is general agreement how benefits
and costs ought to be measured. However, when programs or changes in environmental
resources are involved there is a possibility of large passive use values which inevitably
triggers an ongoing debate among economist and policymakers.

In order to properly judge the appropriateness of changes in environmental
resources a full benefit cost analysis must be employed. This implies that the total (both
direct and passive use) value of the change must be measured as well as the total cost.
Direct use value and its measurement has been well understood. It is clear to most
observers that if an individual absorbs an economic cost in order to make use of an
environmental good then it has value. For example, imagine a hiker who drives to an area
to enjoy the natural scenery and hiking trails. This value can be measured using reveled
preference methods and is rarely controversial. However, things are less clear when one
considers whether that same hiker places a value on a natural resource that is never
directly visited. Such value will leave no behavioral trace from which economists may
infer value. These indirect use values are often referred to as passive use values.

The Contingent Valuation method is the only nonmarket valuation technique able
to measure the passive use component of total value. The method has been widely used,
debated, critiqued and credited over several decades; it is by far the most popular stated preference method used to measure the total value associated with changes in natural resources [7, 8]. Without CV, passive use values could not be directly measured and would only be included in official benefit cost decision-making through the political process or through expert consideration. Ignoring such values would almost certainly tilt policy away from preservation and towards a reduction in environmental goods as benefits would be systematically underestimated.

Estimating the benefits associated with changes in a natural resource is done using one of two common welfare measures, Willingness to Accept (WTA) or Willingness to Pay (WTP). The proper measure may be defined by WTA or WTP depending on the property rights. However, estimates of individual WTA have been difficult to reliably measure [24] (Brookshire, Randall Stoll 1980). As a consequence, WTP is often used as a proxy for WTA [25, 23]. Despite theoretical predictions research suggests that WTA is systematically larger than WTP particularly for nonmarket goods perhaps reflecting preference uncertainty. Recent experimental methodology and field experiments have demonstrated that these differences can be attenuated with proper specification but there is yet to be a solid understanding of the determinants of the disparity and thus it still plagues researchers [16, 21]. The paired comparison method has been used to measure WTA in a theoretically consistent way and has been shown to provide reasonable estimates [10, 17, 20].

Government agencies began to seek monies in order to recover damages created by lost use and nonuse values among environmental resources in the 1980’s. Examples of such instances are the Comprehensive Environmental Response, Compensation and Liability Act of 1980 more commonly known as the Superfund Law. This sought to recover claims made based on damaged land that posed a health risk. In order to clean up such sites the government sought to charge those responsible for the clean up. Shortly thereafter the Exxon Valdez oil spill caused severe damage to Prince William Sound, Alaska in 1989 and in passing the Oil Pollution Act of 1990 congress explicitly charged the National Oceanic and Atmospheric Administration (NOAA) with determining and estimating the damages associated with loss use and passive use values. In order to determine whether passive use values can be accurately measured using CVM, NOAA
established a panel of experts to consider its use. The panel recommended the use of the CVM with caveats imposing significant structure onto the method to insure accurate measurement [2].

Major components of a proper CV study include the use of personal interviews, referendum or closed ended question formats rather than open ended formats, accurate and specific descriptions of the policy to be valued, plausibility of the payment mechanism in order create incentive compatibility and reminders of budget and substitute public goods [2, 22, 9, 13, 19].

Despite the detail of the NOAA panel recommendation there is a great variety of CV methods and little sign that agreement has been met. Personal interviews are rarely used in CV as they are thought to introduce a particular bias known as “yea saying” implying that the respondents respond yes to bid levels higher than they otherwise would in order to please the interviewer. The idea of closed ended question formats proved popular as open-ended questions produced many zeros referred to as “protest votes” [18]. Closed ended question formats can take many forms but they all propose dollar amounts that respondents can then reply yes, no or some statement in between (i.e. maybe yes or maybe no etc) to. Closed-ended formats are theoretically incentive compatible and replicate market experience more closely then the name your own price aspect of open-ended formats.

**Criticism of the Contingent Valuation Method**

The first substantial criticism of the CVM is philosophical. Can a hypothetical market for nonmarket goods accurately reveal preferences? CV depends upon the creation of markets within which these nonmarket goods are traded. They then ask, contingent on this market existing what is respondent value. However carefully designed, the question remains hypothetical. Researchers are skeptical whether a hypothetical format can ever reveal accurate measures of value. This general concern is known as hypothetical bias. Schulze et al. report that hypothetical bias seems to be a relevant criticism of the CVM. Note that familiarity and experience appear to be important factors eliminating hypothetical bias [11].

It has been shown that explicit budgetary scripts and reminder of substitutes are ineffective to reduce hypothetical bias (Loomis et al 94 and Neal 95). Therefore, in order to reduce hypothetical bias researchers have taken two approaches. One tries to calibrate WTP estimates for hypothetical bias (Shogren 2001). The suggested rule of thumb is to follow NOAA guidelines and then divide by two. While the other tries to reduce this bias by explaining what it is, suggesting to respondents that often estimates or expressions of value in CV appear too high to be reliable. This method is known as “cheap talk” and has been shown to effectively reduce WTP estimates (Cummings Taylor; Champ). The concern here is that the respondent is responding to the format and attempting to give the “correct” answer rather than expressing an underlying true preference for the environmental good.

Beyond hypothetical bias there are several established biases associated with the CVM. A typical close-ended question format was a Single Bounded Dichotomous Choice (SBDC) in which the respondent is presented with a single monetary amount and asked whether they would support the change if it cost them a certain dollar amount. This format requires a significant amount of observations to estimate a WTP distribution effectively as each observation only signals that the individuals WTP is greater or less than the bid level. To alleviate this, researchers began collecting multiple observations within an individual using a number of formats including iterative bidding and payment cards.

Each of these formats is susceptible to anchoring and starting point bias suggesting that the bid levels, or dollar amounts presented, unduly influence estimated welfare measures. All else equal research implies that higher values will be measured when higher bid levels are employed (Herriges; Shogren and Hayes; Kramer and Kanninen; Bolye et al 97). Further it has been shown that methods using iterative bidding1 produce different estimated WTP distribution estimated from the first and second bid (Boyle Bishop Welsh 85; Desvousges Smith and Fisher; Thayer). The so called embedding effect implies that welfare estimates from CV are unresponsive to the size of the resource improvement so that cleaning a single lake and cleaning 5 lakes will

---

1 In iterative bidding formats respondents answer yes or no to the first bid level and then are asked a second depending on that answer such that if they responded yes the bid is increased while if they answered no the bid is lowered.
be similarly valued [14]. This has been referred to as the scope test and it has been shown that in properly specified CV surveys respondents are sensitive to scope (Carson 97).

CV practitioners often suggest that these anomalies are the result of poorly designed CVM. Preferences are difficult for respondents to uncover and unless they are motivated to do so they are unexpected to. Such biases are unexpected with market goods as the individual must interact in a market which punishes irrationality and rewards rationality. Anomalies, or irrational choices, within the CVM undermine its usefulness as a valuation methodology (Hanley and Shogren 2005). It is the job of the CV design to motivate this effort. Familiarity and experience are thought to help respondents express preferences consistent with economists’ predictions. Plott (1996) has suggested that preferences do exist but need to be uncovered through a process of discovery. List (2005) suggests choice experiments similar to paired comparison experiments offer similar experience and choice context.

The empirical reality that different elicitation methods consistently and predictably estimate opposing welfare measures still confounds researchers who believe that respondents are acting on consistent, rational underlying preferences in contingent markets. If respondents do not act on such preferences then it is difficult to argue that CV measures what it intends [12]. If not then preferences may simply be transient realizations of the context, framing and instantaneous impressions of the choice environment (Tversky and Simonson 1993). This valuation process is shaped by task and context including the payment vehicle, survey instrument, choice set, bid range, starting point, and framing. This concept fits the constructed preference view that suggests preferences are constructed at the time of valuation rather than simply revealed [19]. It has been shown that initial preferences appear constructed or malleable to context while they also appear consistent in subsequent choices, a process known as coherent arbitrariness [1].

Desvousges et al. (1993) suggest that familiarity is a necessary element within the CV survey in order for a respondent to accurately express preference. Familiarity with the goods must be established through the design on the CV survey instrument. This paper suggests that the paired comparison method offers respondents the opportunity to gain familiarity with the policies of interest in a familiar market where choices are made between dissimilar items. This rich context for choices may enable respondents to
discover there underlying preferences that are unavailable in other CV methods. Indeed, as respondents progress through a PC experiments they appear to uncover consistent preferences.

**Paired Comparison Methodology**

The paired comparison method has successfully been used to measure WTA in several valuation studies [17, 20, 10]. Peterson and Brown (1998) calculate a mean willingness to accept (WTA) for each item by bracketing items between dollar amounts using each items' preference score (discussed below). Champ and Loomis (1998), using similar methods, found the WTA inferred from paired comparison data to be robust to context and scope effects. Specifically they find no statistical difference in measured WTA for a public good in two distinctly different choice sets. However, there was a statistical difference in measured WTA when the scope of the public good was altered. Champ and Loomis (1998) conclude that the method of paired comparisons appears to be a promising approach to elicit measures of WTA.

Loomis et al. (1998) develop an interval model which is able to control for important covariates such as income, age and education using a model designed for paired comparison data which measures WTA similar to models using payment card data [5]. The model presented by Loomis et al. requires two necessary restrictions on the data used. First it reduces each individual to a single observation, the bracket within which their valuation exists. Second, it cannot accommodate the choices that individuals make between dissimilar items. The random utility model presented here relaxes these two restrictions and offers practitioners a flexible econometric framework within which to evaluate PC data. Each choice made by an individual will be modeled and all choices, including those made between items will be incorporated into the random utility model.

Recent research investigating PC data has shown that individual choices become increasingly consistent as they make additional choices within the PC so that later choices appear more consistent then early choices [3]. Further this increasing choice consistency observed by Brown et al. (2008) is accompanied by a significant reduction in the error variance of a random utility model fit to the paired comparison data. This has been interpreted as preference learning and implies that the data become less noisy over
choice occasions indicating that respondents are better able to discriminate between items in later choices [15].

These findings suggest that even hypothetical market experience provided through simple paired comparisons may affect respondents’ choices and those nonmarket valuation techniques that rely on only one or a few responses may not be obtaining well-founded values. This finding is in line with the recent report of Bateman et al. (2008) that respondents to a dichotomous-choice contingent valuation survey require repetition and experience with the choice task in order to express preferences consistent with economic theory.

This paper reanalyzes the paired comparison data collected by Peterson and Brown (1998). In the Peterson and Brown experiment all items were economic gains. Respondents were instructed to choose the item in each pair they would prefer if they could have either at no cost. The paired choices were drawn from a set 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) and 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) along with 11 monetary amounts. See Table 1 for details.

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A meal at a restaurant of the respondents' choice not to exceed $15. Meal</td>
</tr>
<tr>
<td>2</td>
<td>A nontransferable $200 gift certificate to a clothing store of the respondents' choice. Clothes</td>
</tr>
<tr>
<td>3</td>
<td>Two tickets and transportation to a cultural or sporting event in Denver estimated at $75. Tickets and Transportation</td>
</tr>
<tr>
<td>4</td>
<td>A nontransferable $500 certificate good for travel on any airline. Airline Tickets</td>
</tr>
<tr>
<td>5</td>
<td>A 2,000 acre wildlife refuge in the mountains west of Fort Collins Colorado purchased by the University. Wildlife Refuge</td>
</tr>
<tr>
<td>6</td>
<td>An agreement among Colorado State University, local business and government to improve the water and air quality in Fort Collins Clean Arrangement</td>
</tr>
<tr>
<td>7</td>
<td>An annual no-cost on-campus weekend music festival open to all students. Spring Festival</td>
</tr>
<tr>
<td>8</td>
<td>A no-fee service providing video tapes of all class lectures in the University library. Video Service</td>
</tr>
<tr>
<td>9</td>
<td>An expansion to the parking garage system on campus so that parking was always easy to find and convenient. Parking Capacity</td>
</tr>
<tr>
<td>10</td>
<td>An expansion of the eating area in the student center. Eating Area</td>
</tr>
</tbody>
</table>

Table 1
Items included in Peterson and Brown 1998

Dollar amounts were 1, 25, 50, 75, 100, 200, 300, 400, 500, 600 and 700.
Items were not paired with themselves and dollar amounts were not compared (it was assumed that larger dollar amounts were preferred). Each respondent made 155 choices, 45 between items and 110 between an item and a dollar amount. For presentation, pairs were randomized across respondent and choice occasion. The pairs were presented on a personal computer and the time respondents took to enter each choice was recorded. Three hundred and thirty students from Colorado State University participated in the study. Four were dropped because of missing data leaving a total of 326 respondents, providing 50,530 individual observations. In addition, the experiment retested 10 consistent and all inconsistent choices within individual after the initial choices were made. The respondents had not been informed that some choices would be repeated, and there was no break in the presentation of pairs to indicate that a new portion of the experiment had begun.

Figure 1: Illustration of Welfare Measures

WTP is measured by BD
WTA is measured by AC

Figure 1 illustrates the choice. The respondent is assumed to begin at point A on indifference curve $U_1$. Offering a choice between an increase in item $Q_0$ to $Q_1$ and $X$, representing a composite market good allows the estimation of the value respondents place on $Q_1$ [17, 20, 10]. The WTA measures the amount of composite good $X$, which we can interpret in dollar terms, required to make the individual as well off at $Q_0$ as at $Q_1$. Here the WTA is represented by AC. Compensating surplus represents what the
individual is willing to give up to retain the increase to $Q_i$ and is measured here as BD referred to as a willingness to pay. If choosing between two items the respondent is assumed to choose the item which obtains the higher indifference curve.

Given a set of $t$ items, the paired comparison method presents them independently in pairs as $(t/2)(t-1)$ discrete binary choices. These choices yield a preference score for each item, which is the number of times the respondent prefers that item to other items in the set. A respondent's vector of preference scores describes the individual's preference order among the items in the choice set, with larger integers indicating more preferred items. In the case of a 21-item choice set, an individual preference score vector with no circular triads contains all 21 integers from 0 through 20. Circular triads (i.e., choices that imply A>B>C>A) cause some integers to appear more than once in the preference score vector, while others disappear.

For a given respondent, a pair’s preference score difference (PSD) is simply the absolute value of the difference between the preference scores of the two items of the pair. This integer, which can range from 0 to 20 for a 21-item choice set, indicates on an ordinal scale the difference in value assigned to the two items.

When a circular triad occurs, it is not unambiguous which choice is the cause of the circularity. This is easily seen by considering a choice set of three items, whose three paired comparisons produce the following circular triad: A>B>C>A. Reversing any one of the three binary choices removes the circularity of preference; selection of the one to label “inconsistent” is arbitrary. However, with more items in the choice set, selection of inconsistent choices, though still imperfect, can be quite accurate. For each respondent, we selected as inconsistent any choice that was contrary to the order of the items in the respondent’s preference score vector, with the condition that the order of items with identical preference scores was necessarily arbitrary. Simulations show that the accuracy of this procedure in correctly identifying inconsistent choices increases rapidly as the PSD increases. In simulations with a set of 21 items and assuming normal dispersion distributions, the accuracy of the procedure rises quickly from 50 percent at a PSD of 0 to
nearly 100 percent at a PSD of 5.³

The included private goods provide the context unique to the paired comparison method. These are common private goods with known value. The choice between two dissimilar goods focuses the respondents attention onto the tradeoffs associated with each alternative rather than on the attributes such as in conjoint analysis. These types of choice are unique among other CV methods. These methods provide little context for choice and force respondents to value these alternatives in isolation rather than in the context of a choice between policy alternatives. The information inferred from choices made across dissimilar items contains valuable information concerning preferences among the set of policy alternatives and need to be placed directly within the random utility model.

Paired comparison data is unique among contingent valuation applications for several reasons. It obtains many observations within individual and the randomness across individuals may attenuate starting point bias and sequence effects. It significantly improves upon the context within which the choice is made. The tradeoffs between alternative public policy alternatives is made explicit and should act as substitute reminders as well as providing policy makers with important rankings of the publics priorities. Further the choices between private goods and money are familiar providing consumers with a benchmark with which to make the more difficult choices between two public goods or a public good and a monetary amount. This choice format is easy as it replicates the closed ended format preferred by economist and it avoids asking for explicit dollar values on public goods. Importantly it allows researchers to retest, within individual, choices deemed to be inconsistent. If, as suggested by previous PC research, respondents become better able to discriminate between choices as the progress through a PC experiment then these retested choices may more accurately reflect underlying preferences. It is well understood that becoming in touch with preferences, particularly for public or unfamiliar goods is a difficult process. Offering consumers the ability to make such tradeoffs may force consumers to focus on these underlying preferences.

The primary shortcoming of the PC data described in this paper is plausibility of payment mechanism and lack of description for the goods. As described, it is a choice

between two objects including monetary amounts, there is no mention of payment. The goods are only described briefly rather than exhaustively in similar CV designs. Both major weaknesses are easily addressed in future work.

This paper will develop novel econometric models in order to analyze PC data within traditional random utility models familiar to most economists. The model allows great flexibility. The paper explicitly takes advantage of the full range of data available within these experiments including the choices made between items which provide data on relative values within bracketed dollar amounts and the retested choices. It also offers guidance for researchers to develop PC data.

**Econometric Specification**

Paired comparison data provide discrete observations on choice data which readily lends itself to estimation using traditional random utility models. Each individual, represented by \(i\) made 155 choices represented by \(j\) while the choice set contained ten items represented by \(k\). Note that the data is set up in rows and columns, as such the item index will be \(k = r, c\). The row contains only the ten items while the column contains the ten items along with the 11 monetary amounts. The monetary amounts represented by \(t_{jck}\) provide the bid level for respondent \(i\) on choice occasion \(j\). Thus,

\[
    u_{ijk} = \alpha_k + \beta_k X_i + \varepsilon_{ijk}
\]

Where

\[
    \varepsilon \sim N(0, \sigma_{\varepsilon_{ijk}}^2)
\]  \hspace{1cm} (0.1)

Where \(u_{ijk}\) represents the unobserved utility associated with item \(k\) for respondent \(i\) on choice occasion \(j\), \(\alpha_k\) represents an alternative specific constant for each item and \(\varepsilon_{ijk}\) is assumed to be mean zero error term that is allowed to be heteroscedastic over items and choice occasions depending on the functional form of the error structure assumed with this variance denoted \(\sigma_{\varepsilon_{ijk}}^2\). By using the variation in \(t_{jck}\) both the location, the mean WTA
for each item, and the scale, the standard deviation of the valuation distribution, are identified [6, 4].

**Paired Comparison Model**

In order to incorporate the choices between items each observation is modeled as an independent choice. This model incorporates all choices made between items. Thus, this data takes advantage of an additional 14625 observations inaccessible with previous econometric models. These choices provide valuable information concerning preferences and are, for the first time, incorporated directly into the choice model. First, consider the choices between items and dollar amounts. The contributions to the likelihood functions take the following form, where \( P_{rc} \), \( P_{cr} \) is the probability that the row (column) item is chosen over the column (row) item. The dependent variable, \( y_{ik} \), equals 0 if the row item is chosen and 1 if the column item is chosen.

\[
P_{rc} = \Pr(u_{ijr} > u_{ijc}) = \Pr\left( \alpha_r + \beta_k X_i + \epsilon_{ijr} > \epsilon_{ijc} \right) = \Pr\left( \epsilon_{ijr} > \left( t_{ijc} - \alpha_r - \beta_k X_i \right) / \sigma_{\epsilon_r} \right) \tag{0.2}
\]
\[
1 - \Phi\left( \left( t_{ijc} - \alpha_r - \beta_k X_i \right) / \sigma_{\epsilon_r} \right) \tag{0.3}
\]

and similarly,

\[
P_{cr} = \Phi\left( \left( t_{ijc} - \alpha_r - \beta_k X_i \right) / \sigma_{\epsilon_r} \right) \tag{0.5}
\]

Now consider the choice between two items

\[
P_{rc} = \Pr(u_{ijr} > u_{ijc}) = \Pr\left( \alpha_r + \beta_k X_i + \epsilon_{ijr} > \alpha_c + \beta_k X_i + \epsilon_{ijc} \right) \tag{0.6}
\]

In order to identify this model the error terms are assumed to be independent but not identical across items. This allows the model to estimate a standard deviation for each item, referred to as an alternative specific error term, while removing any covariance between items. Thus, the probability that the row item is chosen is as follows.
\[ P_{rc} = \left( \frac{\alpha_r - \alpha_c}{\left( \sigma_{\epsilon_r}^2 + \sigma_{\epsilon_c}^2 \right)^{1/2}} \right) \]  

(0.7)

Where \( \left( \sigma_{\epsilon_r}^2 + \sigma_{\epsilon_c}^2 \right)^{1/2} \) is the standard deviation of \( \epsilon_r - \epsilon_c \). Again a similar statement can be made for \( P_{cr} \) and the likelihood function can be constructed as follows.

\[
L(y_{ijk}; \alpha_k, \sigma_{\epsilon_{jk}}) = \prod_i \prod_j P_{rc}^{1-y_{ijk}} P_{cr}^{y_{ijk}}
\]

(0.8)

Note that \( P_{rc} (P_{cr}) \) now represents the probability that the row (column) item is chosen in both the choices between items and the choices between an item and a dollar amount. Therefore there are 4 distinct contributions to the likelihood function.

**Results**

From the base model specified above several variations are considered. There are three demographic variables available which are gender, age and education in years.

The first important variation is the error structure. The simplest model would hold the standard deviation constant across all individuals, items and choice occasions so that \( \sigma_{\epsilon_{jk}} = \lambda \). This model is inconsistent with economic intuition and is unnecessarily rigid. Rather than assume this error structure three variations are estimated and the preferred model is presented in Tables 3 and 4. The first model allows the error structure to vary across items only so that \( \sigma_{\epsilon_{jk}} = \lambda_j \). Previous research suggests that the error structure is decreasing in choice occasion such that the noise in the data is reduced as respondents’ progress through the paired comparison experiment. To accommodate this both linear \( \sigma_{\epsilon_{jk}} = \lambda_j + \beta \cdot j \) and a nonlinear \( \sigma_{\epsilon_{jk}} = \lambda_j + \beta \cdot 1/j \) error structures are estimated. Here the best fit, given by the AIC statistic is given by the nonlinear error structure which is consistent with previous work. It should be noted that attempts were made to allow the error structure to vary across individuals as well but the model proved unidentifiable.
Lastly one additional variation is estimated that allows the coefficients on the demographics to vary across items or to be constant across items. If it is assumed that a single demographic coefficient works across all items then we implicitly assume that an increase in income will have a similar effect on the WTA of all items in the choice set. Such an assumption would be against economic intuition and the paired comparison data allows us to relax this assumption. Using the AIC statistic it is shown that a model allowing the demographic coefficients to vary across item is preferred and the results are presented in Tables 2 and 3. This paper takes advantage of the retested data and estimates WTA using the preferences implied by these choices. If an originally inconsistent choice is subsequently switched upon retrial the choice is reversed in the estimation.

<table>
<thead>
<tr>
<th>Single Beta, Nonlinear Variance, Original Consistent Choices, Dependent Demographics</th>
<th>ASC</th>
<th>Age</th>
<th>Gender</th>
<th>Education</th>
<th>WTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video Service</td>
<td>192.41</td>
<td>.058</td>
<td>46.72</td>
<td>-6.43</td>
<td>156.70</td>
</tr>
<tr>
<td></td>
<td>(62.16)**</td>
<td>(2.44)</td>
<td>(13.66)**</td>
<td>(5.4)</td>
<td></td>
</tr>
<tr>
<td>Parking Capacity</td>
<td>149.98</td>
<td>.59</td>
<td>39.87</td>
<td>-4.87</td>
<td>138.34</td>
</tr>
<tr>
<td></td>
<td>(56.8)**</td>
<td>(2.33)</td>
<td>(12.82)**</td>
<td>(4.76)</td>
<td></td>
</tr>
<tr>
<td>Meal</td>
<td>23.29</td>
<td>-.85</td>
<td>2.93</td>
<td>.85</td>
<td>20.27</td>
</tr>
<tr>
<td></td>
<td>(9.98)**</td>
<td>(.41)**</td>
<td>(2.18)</td>
<td>(.81)</td>
<td></td>
</tr>
<tr>
<td>Clothing</td>
<td>303.34</td>
<td>-5.84</td>
<td>34.09</td>
<td>-.81</td>
<td>210.10</td>
</tr>
<tr>
<td></td>
<td>(33.48)**</td>
<td>(1.2)**</td>
<td>(7.48)**</td>
<td>(2.73)</td>
<td></td>
</tr>
<tr>
<td>Wildlife Refuge</td>
<td>702.75</td>
<td>-3.15</td>
<td>62.77</td>
<td>-11.04</td>
<td>559.00</td>
</tr>
<tr>
<td></td>
<td>(91.39)**</td>
<td>(3.36)</td>
<td>(20.39)**</td>
<td>(7.46)</td>
<td></td>
</tr>
<tr>
<td>Spring Festival</td>
<td>-5.53</td>
<td>-3.64</td>
<td>-2.23</td>
<td>11.31</td>
<td>66.47</td>
</tr>
<tr>
<td></td>
<td>(51.99)</td>
<td>(2.19)</td>
<td>(11.7)</td>
<td>(4.32)**</td>
<td></td>
</tr>
<tr>
<td>Tickets</td>
<td>97.81</td>
<td>2.63</td>
<td>30.12</td>
<td>-1.68</td>
<td>158.69</td>
</tr>
<tr>
<td></td>
<td>(34.28)**</td>
<td>(1.33)**</td>
<td>(7.57)**</td>
<td>(2.96)</td>
<td></td>
</tr>
<tr>
<td>Eating Area</td>
<td>-59.24</td>
<td>3.02</td>
<td>23.78</td>
<td>-.099</td>
<td>23.65</td>
</tr>
<tr>
<td></td>
<td>(29.99)**</td>
<td>(1.11)**</td>
<td>(6.59)**</td>
<td>(2.53)</td>
<td></td>
</tr>
<tr>
<td>Travel</td>
<td>537.68</td>
<td>5.75</td>
<td>81.12</td>
<td>-4.42</td>
<td>676.34</td>
</tr>
<tr>
<td></td>
<td>(52.67)**</td>
<td>(1.82)**</td>
<td>(11.62)**</td>
<td>(4.49)</td>
<td></td>
</tr>
<tr>
<td>Air Agreement</td>
<td>193.28</td>
<td>10.77</td>
<td>114.45</td>
<td>-7.7</td>
<td>423.03</td>
</tr>
<tr>
<td></td>
<td>(79.29)**</td>
<td>(3.49)**</td>
<td>(18.12)**</td>
<td>(6.59)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2

Estimated WTA for each item assuming average Age and Education and Male

17
### Conclusion

The paired comparison method offers policymakers an alternative CV method which measures WTA and offers several advantages over traditional CV methods. Multiple observations within individual and a rich context of choice allow respondents to gain familiarity and experience with the choice set and the public policy alternatives. Previous research suggests that such experience will enable respondents to reveal consistent preferences. Although beyond the scope of the current paper future work within PCCV must determine the impact on valuation experience within a paired comparison experiment has and, importantly, whether if effects the valuation in a way consistent with expectations.
Bibliography


