DISSERTATION

CRITERION VALIDITY TESTS OF THE CONTINGENT VAUATION METHODS WITH MARKET/NON-MARKET COMPARISONS OF HYPOTHETICAL BIAS AND WILLINGNESS TO PAY ESTIMATES COMPARED WITH PAIRED COMPARISON ESTIMATES OF WILLINGNESS TO ACCEPT

Submitted by

Beatrice Lucero

Department of Agricultural and Natural Resource Economics

In partial fulfillment of the requirements

For the Degree of Doctorate of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2003

COLORADO STATE UNIVERSITY

July 14, 2003

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY BEATRICE LUCERO ENTITLED CRITERION VALIDITY TESTS OF THE CONTINGENT VALUATION METHODS WITH MARKET/NON-MARKET COMPARISONS OF HYPOTHETICAL BIAS AND WILINGNESS TO PAY ESTIMATES COMPARED WITH PAIRED COMPARISON ESTIMATES OF WILLINGNESS TO ACCEPT BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTORATE OF PHILOSOPHY.

Committee on Graduate Work

Adviser

Acting Department Chair

ABSTRACT OF DISSERTATION

CRITERION VALIDITY TESTS OF THE CONTINGENT VALUATION METHODS WITH MARKET/NON-MARKET COMPARISONS OF HYPOTHETICAL BIAS AND WILINGNESS TO PAY ESTIMATES COMPARED WITH PAIRED COMPARISON ESTIMATES OF WILLINGNESS TO ACCEPT

This dissertation investigates the sources of hypothetical bias in contingent valuation methods and differences between willingness to pay (WTP) and willingness to accept (WTA) measures. It involves two separate experiments including six independent treatments all of which utilize the same private good (art print). Hypothetical and actual cash WTP are elicited with both open-ended and dichotomous choice (DC) question formats, which are compared for differences in hypothetical bias. The effectiveness of a *reminder* statement in eliminating or reducing hypothetical bias in both question formats is tested. The *reminder* requested that respondents not report what they thought a fair market price was for the good and to act as if they were in a real market. The DC data was tested for starting point bias. WTP measures elicited with the DC question format were compared with WTA measures for the same art print elicited with the method of paired comparison, which involves having individuals make binary choices between receiving a particular good or a sum of money and inferring WTA from the ranking of dollar amounts and the good of interest. WTP and hypothetical bias are compared for those in the market and those not in the market for art prints.

With the open-ended question format, equality of hypothetical and actual WTP was rejected by the Multi-Response Permutation Procedures at the 0.05 level of significance when the *reminder* was included, but it was not rejected by the Kolmogorov-Smirnov test of distributions, and the Mann-Whitney test of medians rejected equality at

the 0.10 level of significance but not at the 0.05 level. Hypothetical WTP was about two times larger than actual cash WTP.

With the DC question format, neither the non-asymptotic approach to the chi-square statistic developed by Mielke and Berry or Fisher's exact test reject equality of hypothetical and actual cash WTP based upon the raw data. However, the Mann-Whitney test of medians rejects equality of the hypothetical and actual cash WTP based upon logit estimated data according to Hanemann's utility difference approach; but a likelihood ratio test and the method of convolutions (both based upon the logit estimated data) reject equality of hypothetical and actual DC responses at the 0.05 but not the 0.01 level. The logit estimated data produced hypothetical WTP measures about two times the estimated actual cash WTP. No starting point bias was found to exist in the estimated data.

Paired comparison estimates of mean (median) WTA for the art print is estimated at $59($52) compared with the DC contingent valuation estimates of mean (median) WTP of $28(28). The difference persisted in spite of the chooser reference point employed by the method of paired comparison, so the findings do not support Kahnemann et al.'s proposal that endowment effects and loss aversion are responsible for differences between WTP and WTA. The chapter concludes that the method of paired comparison does not produce valid measures of WTA because it does not offer respondents an option of indifference between goods, which is required by the microeconomic axiom of completeness.

The market/non-market analysis of the independent samples indicate that hypothetical WTP, both with and with no *reminder*, and actual WTP are all significantly greater for those in the market for art prints than for those not in the market. The data

indicate that those in the market for art prints and those not in the market come from two separate and distinct populations with separate and distinct WTP behaviors, and it is statistically inappropriate to combine them in one sample. Market/non-market comparisons indicate that the *reminder* was effective in eliminating any significant hypothetical bias for those in the market for art prints and those not in the market for art prints, but it was not effective in eliminating hypothetical bias for those who neither agreed nor disagreed that they were in the market for art prints.

Those not in the market caused skewness in our samples, which were non-normally distributed. Their bids consisted of many zero and low bids. Thus non-parametric statistics are analyzed in depth, and unique applications are employed.

Beatrice Lucero
Department of Agricultural and Natural Resource Economics
Colorado State University
Fort Collins, CO 80523
Summer 2003

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

THE NEED FOR A MEASURE OF THE TREADOFFS INVOLVED IN ALLOCATING LIFE-SUSTAINING ENVIRONMENTAL RESOURCES FOR ONE USE OVER ANOTHER

In a world with an ever-increasing population and demand for life-sustaining environmental resources, it has become more important than ever to recognize and measure the tradeoffs involved in using them for one use over another. The extreme difficulty of this task lies not only in identifying and taking into account all of the ecological benefits and tradeoffs of any given component of environmental quality, given the complexity and dynamics of ecosystems, but also in the fact that the tradeoffs involving these resources cannot be measured as they are in markets for private goods because of the non-excludability and non-rivalry characteristics of pure public goods. There is nevertheless a need for some baseline monetary measure of these tradeoffs, albeit imperfect, for resources are limited and even governments must make tradeoffs as they operate within their budgets.

The contingent valuation method (CVM or CV) is a survey tool that economists use to elicit monetary values that individuals place on public goods, including life-sustaining environmental resources, usually in the form of their stated willingness to pay (WTP). It is not perfect. In fact, it has numerous methodological problems, and its reliability and validity have been challenged from numerous directions. The method may not even have face validity, for it assumes away the very problems that cause markets for public goods to fail in the first place. The referendum method of payment was developed to overcome this problem, but it misses the point. The point is not that everyone should pay through additional taxation, but rather that everyone should pay from that part of their budged that they already allocated to the purchase of public goods this fiscal year.

This dissertation reports the findings of two separate experiments involving six independent treatments designed to investigate several of the issues presently challenging CVM's reliability and validity, including hypothetical bias, question formats, starting point bias, and unexplained divergences between WTP and willingness to accept (WTA). It also reports an independent analysis of my proposal that WTP and hypothetical bias are strongly affected by the extent to which respondents are in the market or not in the market for private goods used in laboratory experiments. Readers who are unfamiliar with CVM or who are interested in a more complete discussion of the problems challenging CVM's reliability and validity are referred to Mitchell and Carson (1989).

HISTORY OF THE CONTINGENT VALUATION METHOD

The first CVM survey was performed in the early 1960s by Robert K. Davis who used questionnaires to estimate the benefits of hunting and outdoor recreation in a Maine backwoods area. From the late 1960s and through the 1970s other economists used Davis' technique to measure the benefits of various recreational amenities, waterfowl hunting rights, reduced congestion from other hikers in a wilderness area, urban parks, pollution control, air visibility benefits, reduced risk of dying from a heart attack, and improved water quality at beaches. In an attempt to legitimize the CV method, several of the early CV studies engaged in construct validity tests, comparing their findings with those of other techniques which had already reached widespread acceptance for valuing recreational use benefits, including travel cost and property value models. Economists have subsequently used CVM to measure the benefits of recreation, hunting, water quality, decreased mortality risk from a nuclear power plant accident, toxic waste dumps, aesthetic benefits from foregoing construction of a geothermal power plant accident,

aesthetic and health benefits of air quality, benefits of the public collection and dissemination of grocery store price information, and benefits of government support for the arts. (Mitchell and Carson:9-14.)

In 1979, the Water Resources Council published its newly revised "Principles and Standards for Water and Related land Resources Planning" in the *Federal Register*. This important document set forth the guidelines for federal participation in project evaluation, which specified those methods that were acceptable for use in determining project benefits. The three recommended methods were CVM, travel cost, and the unit day value method.

> [T]he U.S. Army Corps of Engineers has begun to use CVM to measure project benefits and by 1986 the various engineer corps districts and the corps' Institute of Water Resources had conducted fifteen to twenty studies of varying degrees of sophistication and had published a handbook on the method for corps personnel. Contingent valuation has also been recognized as an approved method for measuring benefits and damages under the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (Superfund), according to the final rule promulgated by the Department of Interior (1986). (Mitchell and Carson:13.)

> Funding from the U.S. Environmental Protection Agency has played a particularly important role in contingent valuation's development. Agency economists recognized early that the prevailing methods of measuring benefits, such as the travel cost method, would be of limited use in valuing the benefits of pollution control regulations they would be responsible for implementing. In the mid 1970s the agency begun to fund a program of research with the avowed methodological purpose of determining the promise and problems of the CV method. At first, almost all of the CV studies funded by this program were designed to test various aspects of the method and to establish its theoretical underpinnings. As the method became better understood, and the agency's mandate to subject proposed regulations to benefit-cost analysis was given sharper focus by the Reagan administration under Executive Order 12291 (Smith 1984), EPA's interest shifted to ascertaining just how effectively contingent valuation can be used for policy purposes. As part of this effort the EPA commissioned a state-of-the art assessment of the CV method in 1983 as a way to step back and reflect on past achievements, remaining problems, and future possibilities. A notable feature of this assessment was the involvement of a review panel of eminent economists and psychologists, including Nobel laureate Kenneth Arrow. During a conference at

> Palo Alto, California, in July 1984, these and other scholars who were actively involved in contingent valuation research offered their views about the method's promise as a means for evaluating environmental goods. The leading authors of the state-of-the art assessment concluded that although the method shows promise, some real challenges remain. (Mitchell and Carson: 13-14.)
>
> [Mitchell and Carson themselves] argue that it is still quite difficult to obtain data from CV surveys which are sufficiently reliable and valid to use for policy purposes, and that the dangers of attempting to draw conclusions from poorly conceived or poorly executed studies are great. Contingent valuation is based on a technique (survey research) with which economists have little expertise, yet it presses this technique close to the limits of its capability. At the same time, the economic theory upon which contingent valuation is based is not well known among sociologists or survey research methodologists, nor is it accepted by them. There is as yet no consensus among CV practitioners about such fundamental questions as what type of market structure they should emulate and what types of errors they should be trying to avoid. (Mitchell and Carson: 15-16.)

PROBLEMS WITH CVM

Hypothetical Bias, Question Formats, and Starting Point Bias

Although there have now been over one thousand applications of CVM (Carson et al. 1994), few of these studies have tested the validity of CVM responses, especially criterion validity, which involves multiple measures of the same concept. One especially important test of criterion validity is a comparison of hypothetical WTP to actual cash payments. Studies in CVM have more recently turned to criterion validity tests of two basic types, field experiments and laboratory experiments.

The field experiments fall into two groups: those of WTP for hunting permits (Bishop and Heberlien 1979; Welsh 1986), and those involving contributions (essentially donations) toward environmental improvements (Seip and Strand 1992; Duffield and Patterson 1992; Champ et al. 1994). The latter set of studies has the advantage of experimenting on public goods, which is an important focus of CVM, but are potentially

subject to free-riding when measuring actual payments because they use a donation payment vehicle.

Lab experiments have relied upon common market goods because they are deliverable and familiar, based on the belief that if CVM cannot get it right for these goods there is much less hope for more complex and unfamiliar environmental services (Cummings et al. 1986). Lab experiments comparing actual cash and hypothetical WTP for private goods have the advantage of careful control of procedures, use of a well-defined good, and avoidance of free riding. However, experimental designs with deliverable goods may be plagued by their own set of problems. Specifically, in stating their WTP in the hypothetical market, some individuals may be stating what they believe a fair price to be or what the actual market price would be, rather than what they would actually pay for the good right then and there. Respondents may also project into a less constrained environment when answering the hypothetical WTP question as compared to the actual cash WTP question, i.e., they might be stating what they would pay after payday, as opposed to taking into consideration only their presently available funds. The field experiments and lab experiments have consistently revealed hypothetical WTP to be significantly greater than actual WTP, so this problem is referred to as "hypothetical bias". I propose that WTP and hypothetical bias are largely affected by the extent to which respondents are in the market for the private good used in the experiment.

Bishop and Heberlien and Welsh also showed that divergences between hypothetical and actual cash WTP varies according to whether an open-ended or dichotomous choice (DC) question format is used. Kealy and Turner (1993) found a statistically significant difference between open-ended and DC mean WTP for their

public good (acid rain prevention) but not for their private good (a chocolate bar). In open-ended CVM, respondents are simply asked to state their WTP for the good as defined in the experiment or survey. In DC CVM, respondents are asked to state affirmatively or negatively whether they would pay the dollar amount specified in their survey, which dollar amount varies randomly from respondent to respondent. Maximum WTP is then statistically inferred (Hanemann 1984).

It has been proposed that perhaps this hypothetical bias is the result of using open-ended question formats; however, it is not yet clear which question format is preferred, as both types of questions formats have their respective problems. The open-ended question format is not considered incentive-compatible because it is believed that participants will not reveal their true maximum WTP so that they can capture some consumer's surplus. It is also different from the normal price taking behavior where consumers react to posted prices and from public good decisions voters make in a referendum (Hoehn and Randall 1987). To better mimic market price-taking behavior, it is now more common to use the DC question format. However, use of the DC format is not without its costs. First, compared to directly asking individuals their WTP, the DC format is statistically inefficient, requiring substantially larger samples for the same level of precision. If in-person interviews are to be used as recommended by the NOAA panel (Arrow et al. 1993), then the cost savings of open-ended questions could be substantial. Second, the DC question format may allow biases unique to the Yes/No format. For example, the DC format might result in symbolic votes in favor of the environmental program, not because the respondent would actually pay the posited price for it, but rather to register their support for providing the public program (Brown et al. 1996). The DC format may also

encourage "yea saying," whereby the posited bid on their survey is accepted as a cue of what a reasonable payment would be (Kanninen 1995; Mitchell and Carson).

Cummings et al. (1995) compared hypothetical and actual cash DC CVM WTP for private goods in three separate experiments with different samples and found that hypothetical DC WTP overstated buying behavior relative to actual cash DC WTP for juicers, calculators, and chocolates. The comparisons were really not comparisons of WTP values, but rather the percentage of Yes responses to the same dollar amount when payment was real versus when payment was hypothetical. Cummings et al. (1995) found the percentage of hypothetical and real dollar percent Yes responses to be significantly different using a chi-square test for all three goods. (His price for the good was usually less than the market price.)

Kealy et al. (1988) compared open-ended and DC CVM WTP using both hypothetical and actual cash transactions for a Cadbury chocolate bar. They found a significant difference between hypothetical and actual cash WTP using the DC question format. They also found no significant difference between open-ended and DC WTP when the transactions were hypothetical; however, they did find a significant difference in the actual cash transactions, with the open-ended estimates of WTP being greater than the DC estimates. These results seem somewhat opposite from what one would have expected. With hypothetical transactions, the question format results were similar, but with the supposed discipline of the cash market they were different. Now even the authors caution about generalizing these results to environmental amenities since candy bars are an often experienced good for which there is a well known market price.

Neill et al. (1994) compared open-ended CVM WTP with actual cash WTP elicited in a second price/Vickery auction. In a second price Vickery auction, the good goes to the highest bidder but at the second highest bid. It is considered to be incentive-compatible because it is believed that respondents will have incentive to reveal their true maximum WTP, knowing that they will have to pay less than their actual bid amount if they are the highest bidder.

The Neill et al. study involved two separate experiments with two different goods, an original painting depicting the desert southwest and a framed map. The respondents were students and were not given a cash endowment, but rather were required to pay "out of pocket" with their own funds, although short-term interest-free loans were available. In the painting experiment, Neill et al. compared hypothetical open-ended CVM WTP with actual cash WTP elicited in a second price Vickery auction. The CVM and Vickery auctions yielded mean WTP estimates of $37 and $9.50, respectively. The distributions underlying these means are significantly different according to Kolmogorov-Smirnov tests. While the CVM was a one-shot case, the students in the Vickery auction were given training sessions and explained the logic underlying the incentives for truthful bidding. This may have introduced some bias against CVM, although the second map experiment was designed to counter this problem.

Neill et al.'s map experiment involved three treatments, a hypothetical open-ended CVM and actual cash and hypothetical second price Vickery auctions, so that they could test whether the divergences between CVM and the Vickery auctions were due to differences between hypothetical and actual cash treatments or differences between second price Vickery auctions and CVM treatments. They found no statistically

significant difference between the two hypothetical treatments (hypothetical open-ended CVM mean WTP of $109 compared with hypothetical second price Vickery auction mean WTP of $110); however, they did find statistically significant differences between these values and the actual cash second price Vickery auction mean WTP ($12). Thus they concluded that the map experiment indicated that the differences found in the painting experiment were also due to hypothetical bias and not lack of incentive-compatibility with the open-ended CVM question format.

The issue of starting point bias arises within the context of bidding games and the DC question format. It occurs when WTP is influenced by a value that has somehow been introduced in the experimental setting. Specifically, respondents may indicate that they would be willing to pay the amount posited in a DC survey not because they would really be willing to pay that amount, but rather because they interpret the posited bid amount as a cue of what they should be willing to pay or of what a reasonable bid amount would be. "Studies have shown that even if a respondent rejects the initial bid, starting points well above the respondent's true WTP amount will tend to increase the revealed WTP amount, while starting points well below it will tend to decrease it." (Mitchell and Carson:100.)

Unexplained Divergences Between WTA and WTP

In addition to the aforementioned tests of criterion validity, economists have sought to test the equality of WTP and willingness to accept (WTA) in the face of seemingly negligible or zero income effects in private goods experiments, pursuant to Willig (1976) and in public goods experiments, pursuant to Randall and Stoll (1980). These tests of equality in both private goods experiments and in CVM have consistently reported ratios of WTA/WTP in the range of 1.6 to 16.6 (Fisher et al. 1988; Cummings et

al. 1986). The large divergence persists even with actual cash experiments, as shown by Welsh (1986), and when goods are actually exchanged, as shown by Knetsch and Sinden (1984). These empirical findings are counter to the equality of WTP and WTA predictions by Willig and Randall and Stoll. Therefore, the empirical results have generally been reported as inconsistent with theoretical predictions.

Kahnemann et al. (1990) explain these large disparities in terms of endowment effects and loss aversion. Horowitz et al. (1996), however, find only limited support for these explanations. With regard to private goods experiments, I believe the divergence arises from the experimental design involving the estimation of WTP and WTA from individuals not likely to be in the market for the private experimental good and WTA for a good individuals received at no cost in the experiment.

At this time, the consensus remains (Mitchell and Carson; Arrow et al.) that current benefit estimating approaches such as CVM do not appear capable of reliably measuring WTA. Thus, the hazard of underestimation of public good benefits remains an important practical issue. A panel appointed by the National Oceanic and Atmospheric Association to evaluate CVM's ability to elicit valid and reliable money measures of welfare changes associated with changes in environmental quality for damage assessment purposes has recommended the use of CVM, but strictly in a WTP question format, because WTP is the more conservative measure in that it is consistently less than WTA (Arrow et al.). This is recommended regardless of whether the appropriate measure of welfare change is WTA or WTP, as determined by property rights.

USING PAIRED COMPARISON TO MEASURE WILLINGNESS TO ACCEPT

The method of paired comparison (PC) may provide an alternative means of eliciting WTA measures of value for environmental quality, while avoiding the problems plaguing CVM measures of WTA. The theory of PC goes back to Fechner (1860) and was formalized by Thurston (1927). David (1988) and Kendall and Gibbons (1990) provide rigorous treatment of the probability theory of comparative judgment that underlies the method. PC can be applied to valuation by specifying a choice set that consists of carefully defined public or private goods (including one or more target goods of special interest) and sums of money.

Figure 1.1 illustrates the relationship between WTP and PC-derived WTA.

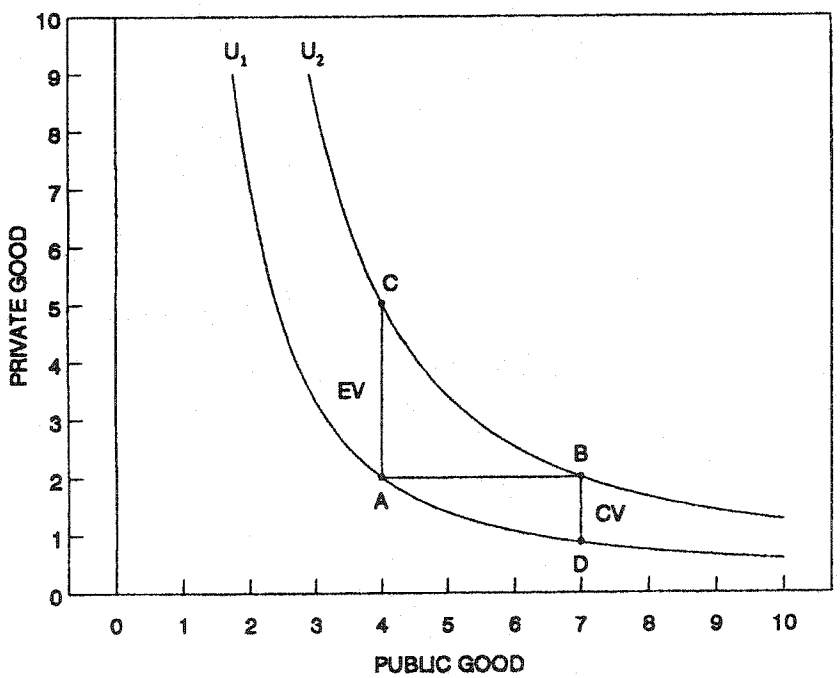


Figure 1.1.

The individual starts at Point A with two units of the private good and four units of the public good. Then the individual is offered a choice between three more units of the private good or three more units of the public good. If the increment in the private

good is chosen over the increment of the public good, the individual's minimum WTA to forego the gain of AB units of the public good is less than the value of the added units of the private good. In Figure 1.1, AC is the minimum amount of the private good that would provide the same level of utility as the addition of AB units of the public good (i.e. Equivalent Variation or EV in Figure 1.1).

Comparing Equivalent Variation (EV) and Compensating Variation (CV)

By asking whether the individual would choose a given increment of the public good or differing amounts of money, the method of PC can bracket WTA within two different dollar amounts. For example, in Figure 1.1, one could identify the point of indifference by asking the individual if they would choose two *more* units of money or three *more* units of the public good and a second choice between four more units of money or three more units of the public good. The "No" response to two more units and the "Yes" response to four more units would allow the analyst to bracket the minimum WTA between two and four units of money.

Figure 1.1 also illustrates WTP for the gain in AB units of the public good. In this case the individual is asked to state the maximum amount of private good they would sacrifice to obtain AB, holding utility constant at the original level (U_1). This is a Compensating Variation measure, an amount equal to BD. In this example, therefore, Compensating Variation is less than Equivalent Variation (BD<CA), which implies that WTA>WTP. However, if endowment effects and loss aversion truly explain the difference between the two measures, then the difference between WTP and WTA in this experiment should not be as large as that measured in past CVM experiments because PC employs a chooser reference point. Kahnemann et al. suggest loss aversion occurs if the

minimum compensation a person would require to move from an existing endowment at point B with seven units of the public good to four units of the public good is greater than AC units of the private good. In Kahnemann et al.'s view, the individual's indifference curve would actually be kinked at the endowment point B and rise more steeply than U_2.

There are at least three conceptual advantages of PC relative to traditional or single bound DC CVM. First, an individual is asked to value one good within the context of a bundle of goods. The number of goods in the bundle and the type of competing goods can be varied by the researcher to make the respondent aware of the policy-relevant tradeoffs. If the government can only provide one or two public goods or services, the choice set could include these possibilities. This should make valuation of the good of interest reflect the presence of the other choices. Most CVM surveys consider just one good, or at most three (Hoehn and Loomis 1993). Second, the repetitive choices between different dollar amounts and the good provide the opportunity to bracket the individual's valuation between a lower and upper dollar amount. This allows for more precise valuation than is possible with a single bound DC CVM (Hanemann et al. 1991), although there is some concern about how to statistically analyze multiple responses from the same individual (Cameron and Quiggin 1994; Alberini 1995). Third, such repeated choices with randomized bid amounts allow for exploring the incidence and nature of intransitive choice.

OBJECTIVES

This dissertation reports the results of two separate laboratory experiments designed to investigate the issues of hypothetical bias, question format, starting point bias, and disparities between WTP and WTA. Both experiments use the same private

good, an unframed art print of a wolf standing in a forest called "*The Shadow Hunter*". The first experiment compares estimates of WTP using both open-ended and DC question formats in both hypothetical and actual cash scenarios. This experiment has five independent treatments, and the primary tests involve comparisons of these independent samples. However, respondents who initially participated in the hypothetical treatments were also given the actual cash treatment immediately after all information had been collected for the hypothetical treatments. Although we recognize that these paired responses are path-dependent, we believe that such a comparison is very informative, and they do allow for a powerful test (Neill et al.). The five independent treatments are listed in Table 1.1. Future references to treatment Nos. herein will correspond to the order reflected in Table 1.1.

Table 1.1.

Independent Treatments in First Experiment

Question Format	**Hypothetical**	**Hypothetical**	**Actual Cash**
Open-Ended	WTP^{OE}(h:no reminder)	WTP^{OE}(h:reminder)	WTP^{OE}(a)
Dichotomous Choice	Not performed	WTP^{DC}(h:reminder)	WTP^{DC}(a)

The test results from the data analyses of all five of these treatments in this first experiment are presented in the two following chapters, with Chapter Two focusing on the open-ended question format and Chapter Three focusing on the DC question format. The findings from the data analyses of the DC question formats are subsequently compared with PC estimates of WTA in Chapter Four. Chapter Five presents market/non-market comparisons of hypothetical and actual WTP for those in the market for art prints and those not in the market for art prints utilizing the same data presented in Chapters Two and Three in order to investigate my proposal that WTP and hypothetical bias are

largely affected by the extent to which respondents are in the market or not in the market for the private goods used in laboratory experiments.

In Chapter Two, WTP from the actual cash and hypothetical open-ended treatments are compared in order to address the following objectives: (1) test whether there is a significant difference between hypothetical and actual cash WTP estimates for a market good with the open-ended question format; (2) test whether some of the disparity between hypothetical and actual cash WTP can be reduced by explicitly instructing respondents in the hypothetical treatment that we do not wish their estimate of a fair market price (This involves the inclusion of a "reminder" statement in one of the hypothetical treatments.); (3) test whether the respondents' being in the market for the subject good is a significant determinant of WTP; and (4) evaluate the robustness of the Neill et al. results using both independent and paired sample responses.

In Chapter Three, WTP from the hypothetical and actual cash DC treatments are compared with each other and with WTP from hypothetical and actual open-ended treatments to address the following objectives: (1) test whether there is a significant difference between hypothetical and actual cash WTP for a private good with the DC question format; (2) compare the performance of the open-ended and DC WTP question formats in hypothetical and actual cash treatments for the same private good; and (3) test for starting point bias in the DC question format.

In Chapter Four, the results of our second experiment are reported. Estimates of WTA for the same art print used in the first experiment are derived through the method of PC. A lower bound on WTA is estimated by asking the respondents to make a binary choice between two alternative gains, for example, $50 or a specified private good. If the

good is selected, then the lower bound for WTA for the private good is greater than $50. These estimates of WTA are compared with $WTP^{DC}(a)$ and $WTP^{DC}(h{:}reminder)$ from Table 1.1 in order to address the following objectives: (1) test whether there is a significant difference between CVM-derived WTP and PC-derived WTA; (2) test whether the chooser reference point employed by PC avoids the loss aversion proposed by Kahnemann et al. because the individual is comparing two alternative gains; and (3) test whether differences between PC-derived WTA and CVM-derived WTP are smaller than the historical gap between CVM-derived measures of WTA and WTP.

In Chapter Five, an independent analysis of my proposal that WTP and hypothetical bias are strongly affected by the extent to which respondents are in the market or not in the market for private experimental goods is presented. This involves a two-part analysis. First, the original samples are divided according to those who either agreed or strongly agreed with the statement that they were in the market for art prints (*MARKET*) and those who neither agreed nor disagreed, disagreed, or strongly disagreed with the statement that they were in the market for art prints (*OTHERS*). These two sub-samples are then compared with the objective of determining whether there are significant differences in *WTP(h:reminder)*, *WTP(h:no reminder)*, and *WTP(a)* between them. These two sub-samples are then further compared in a re-testing of null hypotheses [2.1] – [2.7] presented in Chapter Two but with the original samples again divided between those respondents in the *MARKET* for art prints and the *OTHERS*.

Then those who neither agreed nor disagreed with the statement that they were in the market for art prints were excluded from both sub-samples, since they really do not belong in either, and since it is really a comparison of those who are strictly in the

MARKET and those who are *"NOT IN THE MARKET"* for art prints that is relevant to market/non-market comparisons of WTP and hypothetical bias. Thus all of the null hypotheses presented in Chapter Five are tested twice – once when WTP behavior for those in the *MARKET* and the *OTHERS* is compared and twice when WTP behavior for those in the MARKET and those *NOT IN THE MARKET* is compared. Conducting both comparisons allows us to identify distinct patterns of WTP and hypothetical bias behavior for all three groups of respondents – those in the market for art prints, those not in the market for art prints, and those who neither agreed nor disagreed that they were in the market for art prints.

The specific hypotheses and the methods used for testing them are discussed in detail in their respective chapters. Overall conclusions are presented in Chapter Six.

CHAPTER TWO

IMPROVING VALIDITY EXPERIMENTS OF CONTINGENT VALUATION METHODS: RESULTS OF EFFORTS TO REDUCE THE DISPARITY OF HYPOTHETICAL AND ACTUAL WILLINGNESS TO PAY[1]

[1] This chapter draws largely upon: Loomis, John, Thomas Brown, Beatrice Lucero, and George Peterson. 1996. "Improving Validity Experiments of Contingent Valuation Methods: Results of Efforts to Reduce the Disparity of Hypothetical and Actual WTP." *Land Economics* 72(4):450-61; however, it may depart significantly in some respects.

TESTING CRITERION VALIDITY OF CONTINGENT VALUATION

This chapter reports the results of comparing the estimates of WTP elicited in the three open-ended treatments reflected in Table 1.1. Since only the open-ended treatments are discussed in this chapter, the superscripts denoting the treatments as open-ended are dropped in this chapter only. Actual cash WTP is compared with hypothetical WTP, elicited both with and without the *reminder*, in order to address the following objectives: (1) test whether there is a significant difference between hypothetical and actual cash WTP estimates for a market good with the open-ended question format; (2) test whether some of the disparity between hypothetical and actual cash WTP can be reduced by explicitly instructing respondents in the hypothetical treatment that we do not wish their estimate of a fair market price (This involves the inclusion of a "*reminder*" statement in one of the hypothetical treatments.); (3) test whether the respondents' being in the market for the subject good is a significant determinant of WTP; and (4) evaluate the robustness of the Neill et al. results using both independent and paired sample responses.

We have acknowledged the proposal that the open-ended WTP question format is not incentive-compatible; however, since we use this same question format to elicit both hypothetical and actual cash WTP, the difference in WTP should be due to the hypothetical nature of CVM. Whatever understatement of "true" WTP might arise from lack of incentive-compatibility should be present in both the hypothetical and actual treatments. Evidence of this can be found in the results of Neill et al.:152) where, as previously noted, the supposedly incentive-compatible second price Vickery auction format yielded hypothetical WTP ($110)

nearly identical to hypothetical CVM WTP ($109) but a significant difference between these two measures and second price Vickery auction actual cash WTP ($12).

RESEARCH DESIGN

Treatments.

The three open-ended treatments used to test the hypotheses addressed in this chapter are as follows:

No. 1: *WTP(h:no reminder)*: hypothetical WTP asked as in a standard CVM survey, using wording similar to Neill et al.

No. 2: *WTP(h:reminder)*: hypothetical WTP asked after subjects are reminded not to give what they think a fair price is, or what it sells for, and to act as if they were in a real market with their real budget.

No. 3: *WTP(a)*: actual WTP in the form of cash, check, or promissory note.

WTP(a) is treated as the standard for criterion validity in this experiment. However, if the open-ended question format is truly not incentive-compatible, the implication would be that *WTP(a)* would not necessarily be the respondents' maximum WTP; it would be a downward biased measure of it.

Null Hypotheses.

The null hypotheses of criterion validity tested in this chapter are:

[2.1] H_0: *WTP(a)=WTP(h:no reminder);* and

[2.2] H_0: *WTP(a)=WTP(h:reminder).*

The null hypotheses of equality of hypothetical and actual cash WTP for the two paired samples in treatments No. 1 and No. 2, respectively are:

[2.3] H_0: $WTP_i(a)=WTP_i$*(h:no reminder);*

[2.4] H_0: $WTP_i(a)=WTP_i$*(h:reminder)*; and

H_0:B_0=0 and H_0:B_1=1 in equations [2.5] and [2.6]:

[2.5] $WTP_i(a)=B_0+B_1[WTP_i(h:no\ reminder)]$;

[2.6] $WTP_i(a)=B_0+B_1[WTP_i(h:reminder)]$.

where i represents the i^{th} respondent, indicating that these are paired samples.

The null hypothesis that we use to test our objective of determining whether efforts to educate respondents in the hypothetical market that we desire their immediate value as if they were in a real market and not some measure of fair price reduces the disparity between hypothetical and actual cash WTP in independent samples, as compared to Neill et al.'s experiment is:

[2.7] H_0:$WTP(h:no\ reminder)=WTP(h:reminder)$.

The null hypotheses used to test the equivalence of valuation behavior in the independent hypothetical and actual cash treatments are:

H_0:$B_0=C_0$; H_0:$B_1=C_1$; and H_0:$B_2=C_2$ in equations generated by OLS regressions of [2.8], [2.9], and [2.10], respectively:

[2.8] $WTP(h:reminder)=B_0+B_1(MARKET)+B_2(SELL)$;

[2.9] $WTP(h:\ no\ reminder)=B_0+B_1(MARKET)+B_2(SELL)$; and

[2.10] $WTP(a)=C_0+C_1(MARKET)+C_2(SELL)$

where *Bs* are coefficients from equations [2.8] or [2.9], depending on the comparison; and where the variable *MARKET* indicates how strongly the respondent agreed or disagreed with the statement that they were in the market for the type of art print shown and *SELL* is a variable representing what respondents believed the market value of the print to be (which was asked after they had reported their WTP). The conceptual model behind equations [2.8]

- [2.10] was that open-ended WTP was a linear function of whether the respondents were in the market (*MARKET*) for art prints, whether they liked the print (*LIKE*), whether they would purchase the print as a gift for someone else (*GIFT*), and what they thought the fair market value of the print was (*SELL*). The *MARKET*, *LIKE*, and *GIFT* variables were highly collinear, and *INCOME* was not a significant determinant of WTP in these open-ended treatments, so the model was reduced to include only *MARKET* and *SELL* as explanatory variables. It makes sense that if individuals are in the *MARKET* for art prints that they *LIKE* art prints and would buy one as a *GIFT* for a friend.

The null hypotheses that we used to test our objective of determining whether the respondents' WTP is significantly influenced by what they think the object sells for rather than just their own personal WTP are:

H_0:$B_2=0$ in equation [2.8] and [2.9]; and

H_0:$C_2=0$ in equation [2.10].

The null hypotheses that we used to test our objective of determining whether the respondents' being in the market for the subject good is a significant determinant of WTP are:

H_0:$B_1=0$ in equations [2.8] and [2.9]; and

H_0:$C_1=0$ in equation [2.10].

EXPERIMENTAL DESIGN

University clerical and administrative staff in academic and nonacademic units were recruited and paid $20 for attending a 45-minute session on campus. The sample sizes ranged

from 30 to 35 people per treatment. The same researchers conducted all of the sessions, following a script that was identical except for treatment effects.

Nature of the Good.

The good chosen for the experiment reflects several desirable features in private goods experiments. First, art prints are infrequently purchased and different prints sell at quite different prices, so people would not be likely to know the market price of any given print. The objective was to minimize the likelihood that the respondent would simply try to state the market price. Second, the good has readily observable characteristics, so there is minimal ambiguity in terms of what the product is. Finally, we wanted a good that is not too expensive so it could be paid for by the highest bidding respondents in the form of cash, the current balance in their checking accounts, or a promissory note payable within three weeks.

We pretested several wildlife art prints on a separate sample of university staff and settled on a signed wildlife print of a wolf standing in a forest called "The Shadow Hunter". We conducted five pretest treatments with different samples of university staff to fine-tune the procedures for the three different treatments and to better understand the thought processes used in both the actual cash and hypothetical market scenarios. Based on post-treatment wrap-up discussions and written comments elicited from these sessions, several revisions were made to procedures and instructions. Changes included adding questions for respondents to rate the prints prior to the auction and changing the wording in the *reminder* to counter tendencies of the respondents in the hypothetical market to not fully consider their current budget. This process continued until respondents' comments after the last sessions indicated they understood the task before them in each treatment as we intended.

Setting of Experiments.

All the treatments were conducted in a classroom with the participants sitting at every other seat to maintain privacy and eliminate any discussion with each other. At the beginning of the session, individuals were shown the art print. They were then asked to respond to a series of statements regarding how well they liked the art print, whether they would buy it for themselves or a friend, and if they were in the market for this type of art print. A standard five-point Likert scale was used, where strongly agree is coded as five, agree as four, neither agree nor disagree as three, disagree as two and strongly disagree as one (Bailey 1987:346). Next, individuals were instructed to read and complete the bid submittal page. When everyone had finished, the sheets were passed forward. The respondents then filled out a sheet on their demographics.

Respondents in treatments No. 1 and No. 2 were first given the *WTP(h:no reminder)* and *WTP(h:reminder)* treatments, respectively. They were then given the *WTP(a)* treatment after all of the documentation necessary to analyze the independent treatments had been collected. Although the respondents' follow-up actual cash WTP responses in these paired samples may be conditional upon their responses in the initial hypothetical treatment, it will nevertheless be informative to see how the hypothetical and actual cash treatments differ from one individual's perspective. The independent samples remain uncontaminated.

Then in all three sessions, the winner was announced and asked to come forward to complete his or her purchase in front of the group, but the winning price was not announced. Individuals were allowed to pay with cash or check or sign a promissory note payable within

three weeks.[2] The option of paying with a promissory note was included to convince the participants that the highest bidder would really be expected to pay his or her bid amount.

Wording of WTP Questionnaires.

Treatment No. 1. The wording of the *WTP(h:no reminder)* questionnaire was:

*1. You are being asked to participate in a **HYPOTHETICAL** sealed bid auction for this print. We would like to know the maximum amount of money you would pay to take this art print with you at the end of this session, if this one art print were actually for sale, and you would have to pay by August 19.*

Now please write down the maximum dollar amount you would be prepared to pay for this art print.

4. I would bid $_____.

5. Please describe what factors you considered when arriving at your dollar amount.

When you are finished please fold up your answer sheet. When everyone is finished, we will ask you to pass them forward.

Treatment No. 2. The wording of the *WTP(h:reminder)* questionnaire was:

*1. You are being asked to participate in a **HYPOTHETICAL** sealed bid auction for this print. We would like to know the maximum amount of money you would pay to take this art print with you at the end of this session, if this one art print were actually for sale.*

At this time in the survey, we are NOT asking what you think the art print might sell for in a store or what you think its fair price is. Rather, we want to know the maximum amount of money that you would honestly be prepared to pay right now to buy the art print you are being shown if you would really be required to pay your bid amount with cash, write a check today, or sign a Promissory Note payable on or before August 19. Please take into consideration your budget and what you can afford to pay. If what you would pay is different from what you judge a fair price

---

[2]The loan option in our experiment was provided by me, a graduate student who assisted in the development and running of the experiment. In the Neill et al. experiment, the loan was provided by a graduate student who was not officially involved in running the experiment. Our promissory note was payable in three weeks, overlapping a payday to insure that participants near term cash balances were not constraining. This extended time period and pay consideration was developed from debriefing responses provided by participants after one of the pretests. Neill et al.'s (149) loan period appears to have been less than one week in duration.

to be, that is OK. We want to know what you would actually be prepared to pay for the art print.

Take a few moments to think about what you honestly would be prepared to pay for this art print if it were being offered for sale to you today and it would go to the highest bidder. Although the question is hypothetical, we want you to answer as if it were for real - as if you were participating in a real sealed-bid auction and would really have to pay your dollar amount if you were the highest bidder.

STOP. Do not answer questions #4 and #5 until instructed to do so.

(Participants in this hypothetical treatment with *reminder* were asked to read the foregoing instructions and then stop and wait for further instructions. When it was apparent that everyone had finished reading the instructions, the interviewer then reiterated the foregoing instructions verbally to the participants. They were then asked to proceed. The questionnaire proceeded as follows):

Now please write down the maximum dollar amount you would be prepared to pay for this art print.

4. I would bid, and would really be prepared to pay, $_____.

5. Please describe what factors you considered when arriving at your dollar amount.

When you are finished please fold up your answer sheet. When everyone is finished, we will ask you to pass them forward.

As can be seen, the *reminder* version attempted to more fully place the individual in the frame of mind of the real market situation, without actually requiring them to pay. This statement was developed after discussions with pretest participants indicated that they were in a different frame of mind when answering the hypothetical WTP questions as compared to a follow-up actual cash scenario. Second, we wanted individuals to report their WTP for

the print rather than attempt to estimate what they thought a reasonable price would be in a store.

Treatment No. 3. The wording of the *WTP(a)* questionnaire was:

*1. As a part of this experiment, we are now going to conduct a **REAL** auction. This art print will be sold to the highest bidder here today. Only one of these prints will be sold at this auction.*

2. After all bids have been collected, the person who is the highest bidder will be announced and he or she will be obligated to purchase the print at his or her bid price. We will accept cash or check (payable to C.S.U.) for your purchase. We understand that you may not have anticipated the need to bring cash or your checkbook with you today, so we will also accept a signed Promissory Note payable on or before August 19.

In any case, the highest bidder will be required to pay his or her bid amount and will then be able to take the art print with him or her at the end of this session.

*3. Please understand you are participating in a **REAL AUCTION.***

Remember that only one of these art prints will be sold at this auction.

You are under no pressure to bid more than you are actually prepared to pay for this art print. However, since the print will go to the highest bidder, it is in your interest to bid the maximum that you are prepared to pay for the art print.

4. Now take a few moments to determine the maximum dollar amount that you are prepared to pay for this art print.

5. What is the most you are prepared to pay for this art print? I bid $_____.

6. If I am the highest bidder I promise to complete the purchase at my bid amount.

_____
Signature

7. Please describe what factors you considered when arriving at your dollar amount.

When you are finished please fold up your answer sheet. When everyone is finished, we will ask you to pass them forward.

It proved somewhat difficult to develop the wording in the actual cash questionnaire such that there would be no doubt in the respondents' mind that the auction was real, and that the highest bidder would really be expected to pay his or her bid amount. I proposed item No. 6, in addition to offering the option of paying for the print with a promissory note, in an effort to place the respondents in the same obligatory frame of mind that one experiences when signing a promissory note at a bank. It was unanimously agreed that the language, signature, and option of a promissory note accomplished the objective.

DISCUSSION OF STATISTICAL TESTS AND DESCRIPTIVE STATISTICS

Statistical tests considered to test hypotheses of equality for the independent sample open-ended treatments include the two sample *t* test of the means; the Multi-Response Permutation Procedures (MRPP); the Mann-Whitney (aka Mann-Whitney-Wilcoxon) two sample rank test of differences in medians; and the Kolmogorov-Smirnov test for differences in distributions.

Use of the two sample *t* cannot be justified because normality of the distributions, both in their original and natural log forms, was rejected by the Jarque-Bera normality test statistic, which is distributed as a chi-square with two degrees of freedom. (Hall et al. 1990). The distributions are largely skewed because of many zero and low WTP responses.

Multi-Response Permutation Procedures.

The MRPP are non-parametric tests developed by Dr. Paul W. Mielke and his colleagues at Colorado State University. These tests make no distributional assumptions, not even asymptotically. The tests take into account the first three moments, mean, variance, and skewness. Two independent samples under comparison need not have the same shape or

variance, or by symmetric. For small samples, probabilities can be calculated exactly by complete enumeration of all possible combinations under the null hypothesis. For large samples, *p* values are based on the Pearson Type III distribution which is totally specified by the skewness value. Skewed distributions are characteristic of permutation distributions and of our experimental data. (Slauson et al. 1991.)

MRPP is permutation (randomization theory) based. Permutations make efficient use of small sample sizes because exact *p* values can be calculated and degrees of freedom are not lost since no parameters are estimated. The test statistic is equal to the sum of the weighted average distance function values for all permutations in a sample. The underlying permutation distribution assigns equal probability to each possible allocation. MRPP allows the investigator to weight the intragroup average distances by either relative degrees of freedom (C=2) or by relative sample size (C=1). C=2 is the weight used by parametric and other non-parametric statistical tests. The basis for MRPP, on the other hand, is permutations. Degrees of freedom are irrelevant to permutations, because no parameters are estimated. If the samples under comparison are of the same size, then the choice of C makes no difference.

Statistical tests that make distributional assumptions are based upon the variance (sum of squared deviations from the mean). MRPP permits the investigator to determine the geometric space in which the data analysis is performed. If V=1, the data analysis takes place in Euclidean space – the space in which data are measured. This space is metric and as such is easily interpretable. In Euclidean space, the sum of absolute differences are used, as opposed to squared Euclidean space (V=2), which minimizes the sum of the squared

differences. V=2 is used by parametric and other non-parametric permutation tests. Mielke contends that the rejection region (and therefore the power) of any test is highly dependent upon the geometric space in which the data is analyzed. "Euclidean distance based statistics have greater power (the probability of rejecting the null hypothesis when it is false) to detect location (central tendency) shifts among skewed distributions than do squared Euclidean distance statistics." [Slauson et al.:3 referencing (Zimmerman et al. 1985, Biondini et al. 1988).] "Power to detect location shifts in symmetric distributions with Euclidean distance statistics is greater than or equal to power with squared Euclidean distance statistics depending on the distributional form." [Slauson:3, referencing (Mielke et al. 1981; Mielke and Berry 1982; Mielke 1984).] Choosing V=2 permits outliers to have a disproportionate effect on statistical results because their absolute distances are squared.

When V=2 and C=2, MRPP is analogous to the two sample *t* test of equality of means, but without the normality or asymptotic normality assumption. The MRPP test statistic for more than two groups and the F statistic are equivalent. When V=1 and C=1, MRPP is a test of equality of medians. This test of medians is performed on the absolute values of the observations under analysis and does not reduce the values to ordinal values, as does the Mann-Whitney test of the medians. MRPP will also emulate the Mann-Whitney test, but with the use of permutations. This test assumes V=2 and C=2 and utilizes ranked (ordinal) data.

Mann-Whitney Test of Medians.

The Mann-Whitney test is often used in the absence of normally distributed data. This test determines whether the medians of two mutually independent random samples are

significantly different. This test pools the data and sorts it by numerical ascending order while maintaining distinction between the samples under comparison. The observations are then ranked in their sorted order by assigning a value of one to the first observation and adding one to each subsequent observation. The difference of the population medians is then estimated by the median of the differences. (Gibbons 1993.) A *p* value is determined which represents the probability that the difference between the ranks was determined randomly. "The asymptotic relative efficiency of this test relative to the Student's *t*-test is .955 for normal distributions, 1.00 for the continuous uniform distribution, and at least .864 for any continuous distribution..." (Gibbons: 38-40).

Many zeros and tied values violate the continuity assumption and thus present problems for the Mann-Whitney test because it replaces quantitative information with qualitative data. The tied values are necessarily assigned different values so that they can be ranked ordinally. Not only does the Mann-Whitney test destroy information contained in the raw data, but nonmetric squared Euclidean distance also replaces the preferred ordinary Euclidean distance, and a weighting function dependent on degrees of freedom (associated with the *t* test's dependence on normality) is used. (Mielke and Berry 2001).

Kolmogorov-Smirnov Test of Distributions.

Another test frequently used in the absence of normally distributed data is the Kolmogorov-Smirnov test for differences in the distributions of a continuous variable. The was the test statistic reported by Neill et al. The Kolmogorov-Smirnov test involves determining whether the cumulative distribution functions are different for at least one point.

Descriptive Statistics.

The continuity requirement for both the Mann-Whitney and Kolmogorov-Smirnov tests poses a critical problem for the validity of applying either these tests to our data, because our data is not continuous. The fact that our data is non-continuous and characterized by large numbers of zero and tied values is reflected in Table 2.1. Although the Bid Ranges are presented in intervals of five, most of the bids were multiples of five. Table 2.1 also reflects the fact that individuals in treatments No. 1 and 2 were first asked their respective hypothetical WTP and then their actual cash WTP, and that the individuals in treatment No. 3 received only the actual cash treatment.

Table 2.1

Number of Bids Received by WTP Class

	Treatment #1		Treatment #2		Treatment #3
Bid Range	**WTP(h:nr)**	**WTP(a)**	**WTP(h:r)**	**WTP(a)**	**WTP(a)**
0	2	9	6	7	3
1-5	5	8	1	9	8
6-10	1	3	6	6	7
11-15	2	3	3	3	3
16-20	5	5	3	0	3
21-25	3	4	1	4	2
26-30	2	1	3	0	3
31-35	2	1	2	1	0
36-40	2	1	1	1	1
41-45	2	0	0	0	0
46-50	4	0	2	1	2
51-55	1	0	1	0	0
56-60	0	0	0	0	0
61-65	1	0	0	0	0
66-70	0	0	0	0	0
71-75	1	0	3	0	0
>75	2	0	1	1	0

With regard to the continuity assumption of the Kolmogorov-Smirnov test, Cade and Richards (2001:3, 6) state that:

> G-sample and goodness-of-fit tests based on empirical coverages (COV) are for univariate comparisons of grouped data similar to the Kolmogorov-Smirnov family of statistics for comparing cumulative distribution functions (Mielke and Yao 1988, 1990). These statistics are appropriate for continuous univariate responses with no or few tied values.
>
> The empirical coverage tests included in BLOSSOM are related to the Kolmogorov-Smirnov family of tests for equality of univariate cumulative distribution functions. One sample goodness-of-fit and g-sample tests exist. The coverage test statistic is based on the spacings between the order statistics. These tests provide another permutation testing alternative to MRPP for univariate continuous data. Unlike, MRPP, the coverage tests are not appropriate when there are many tied values, as this violates the continuity assumption.

Further, with regard to the problems that non-continuous data characterized by many zeros and tied values pose for the Mann-Whitney test, Snedecor and Cochran (1967: 129-132) state that the significance probabilities are changed, and the power of the test in finding a significant result when the null hypothesis is false is altered.

Given that the distributions associated with our data are non-normal, non-symmetric, non-continuous, and highly skewed, and that we have small sample sizes, MRPP is the most appropriate and powerful test applicable to the analysis of our data. Although it is recognized that our data does not meet the continuity assumptions of the Mann-Whitney and the Kolmogorov-Smirnov tests, the results of these two tests are presented in this chapter as a point of comparison and discussion only because they were reported in Loomis et al. (1996).

Regarding the data presented in Table 2.1, Hypothetical WTP ranged up to $400 in treatment No. 1 and to $100 in treatment No. 2. Actual WTP ranged up to $40, $100 and $50 in treatments No. 1, 2 and 3, respectively. In treatment No. 1, only 7 of the 35 subjects bid the same amount in the hypothetical and actual treatments; and in treatment No. 2, 14 of the 33 subjects bid the same amount in both treatments. In treatment No. 2, the same person who

bid $100 in the hypothetical treatment also bid $100 in the actual cash follow-up treatment and in fact paid this amount; however, she returned the art print that same afternoon and requested her money back, which was returned to her.[3] The fact that the woman who bid this same amount in the actual cash follow-up treatment, even though she was not *really* willing to pay $100 for the print, may be reflective of path-dependency in the paired sample treatments. Perhaps she felt as though she should bid the same amount in both treatments so as to avoid appearing as though she had not bidden truthfully in the hypothetical treatment.

Permutation Tests for Matched Pairs.

Tests considered for comparing our paired samples include the sign test of the medians and the paired sample permutation test equivalent of MRPP called the permutation test for matched pairs (PTMP). The sign test considers only whether observations are greater than or less than the median but does not take into account magnitudes; thus it is not a powerful test. The asymptotic relative efficiency of the sign test relative to the paired *t* test is .67 for normal distributions and at least .33 for any continuous symmetric distribution. (Gibbons:25.) Although our data does not meet the sign test's continuity requirement it was reported in Loomis et al. (1996) and is reported here as a point of comparison and discussion.

As with MRPP, the PTMP can be performed using the squared differences between responses (V=2) and lineal differences between paired responses (V=1). PTMP (V=2) is equivalent to MRPP (V=2), the test of mean equality for non-normal distributions derived from independent samples discussed in detail above. PTMP (V=1) is equivalent to MRPP (V=1), the test of median equality for non-normal distributions derived from independent samples also discussed in detail above. PTMP is the most appropriate and powerful test

[3] This information was not reported in Loomis et al. (1996).

applicable to our paired sample data for the same reasons MRPP is the most appropriate and powerful test for our mutually independent sample comparisons.

Wald Test.

Kennedy (1992:61) recommends use of the Wald test for the significance of equation coefficients in the case of non-normally distributed residuals, as is the case for the residuals generated by OLS regression of equations [2.5] and [2.6]. The Wald test has an unknown small-sample distribution but is asymptotically distributed as a chi-square with degrees of freedom equal to the number of restrictions. The Wald test is also used to evaluate the coefficients generated by OLS regressions of equations [2.8], [2.9], and [2.10], as the residuals from these three equations are also non-normally distributed.

Log Likelihood Ratio Test.

The log likelihood ratio (LLR) test is used to test for the equality of these coefficients. The hypothetical WTP response data and actual cash response data are combined and one regression estimated. If the coefficients are equal, the log likelihood value of this pooled model should not be significantly different from the sum of the log likelihood values of the two independently estimated equations. The LLR test uses a chi-square distribution with degrees of freedom equal to the sum of coefficients in the individual regressions minus the number of coefficients in the pooled regression.

Comparison of Demographics Across Sessions.

Before establishing that any differences between treatments were due to differences in stimulus, it was first determined that the respective samples were not statistically different in terms of standard demographics. To test for this across our sessions we performed one-

way ANOVA's for education (F=1.88, p=0.16), age (F=1.4, p=0.25) and income (F=0.21, p=0.81), which showed the samples are not statistically different at the 0.05 significance level. The *WTP(a)* and *WTP(h:no reminder)* treatments consisted of about 70-80 percent women, but the *WTP(h:reminder)* session included only one male. The sample population of university clerical and administrative staff appears to be over-represented by females. Multiple regression based testing for gender indicates that it is not a significant determinant of hypothetical or actual WTP, so we do not consider this problematic.

Table 2.2 reports mean and median WTPs for our three independent treatments. The large differences between the respective means and medians are indicative of non-normally distributed data.

Table 2.2.

Comparison of Hypothetical and Actual WTP

Treatment	Sample Size	Mean (Median) WTP by Treatment [S.E. of Mean] *Hypothetical no reminder*	*Hypothetical with reminder*	*Actual*	*WTP(h) / WTP(a)*
No. 1	35	$42.34 (25) [11.38]		$11.63 (6) [1.92]	3.64
No. 2	33		$26.29 (20) [4.45]	$13.42 (6) [3.42]	1.96
No. 3	32			$14.48 (10) [2.27]	

RESULTS OF FORMAL HYPOTHESES TESTS

Criterion Validity Tests with Independent Samples.

Table 2.3.

Probability Levels Associated with Testing the Equality of Actual and Hypothetical WTP (With and Without Reminder) From Independent Samples

Hypothesis for Independent Samples	Mann-Whitney (medians)	MRPP (test of means)	MRPP (test of medians)	Kolmogorov-Smirnov (distribution)
WTP(h:no reminder)=WTA(a)	p=0.0017	p=0.0015	p=0.0037	p=0.009
WTP(h:reminder)=WTP(a)	p=0.09	p=0.0216	p=0.0327	p=0.318
WTP(h:reminder)=WTP(h:no reminder)	p=0.21	p=0.218	p=0.314	p=0.511

H_0:WTP(h:no reminder=WTP(a)).

As shown in the first row of Table 2.3, all four test statistics reject equality of *WTP(h:no reminder)* and *WTP(a)* at the .01 level of significance. Different MRPP values for C=1 and C=2 are not reported because our sample sizes are similar, and so the value of C makes no difference.

H_0:WTP(h:reminder)=WTP(a).

The tests reported in the second row of Table 2.3 are mixed. The MRPP tests of the means and medians reject equality at the .02 and .03 levels of significance, respectively. However, the Kolmogorov-Smirnov test does not reject equality, and the Mann-Whitney test of the medians rejects equality at the .10 level of significance but not at the .05 level.

Testing the Effectiveness of the Reminder Statement with Independent Samples.

H_0:WTP(h:reminder)=WTP(h:no reminder).

Our first test is simply whether *WTP(h:reminder)=WTP(h:no reminder)*, where the *no reminder* version mimics a typical CVM survey and the experiment of Neill et al. (1994).

As can be seen in the third line of Table 2.3, all four statistical tests indicate that we cannot reject the null hypothesis. Thus, even though mean and median hypothetical WTP are lower when the *reminder* is used, they are not significantly lower.

Due to non-normality in the residuals and dependent variables, the Wald test is used to test for coefficient significance. Since this restriction is linear, the *F* statistics are valid. (Kennedy.) OLS regressions of equations [2.8], [2.9], and [2.10] generate equations [2.11], [2.12], and [2.13], respectively.

[2.11] *WTP(h:reminder)*=-5.578+8.886(*MARKET*)+.085(*SELL*)
(*F* statistic) (.3131) (7.145) (3.227)
(*p* value) (.5802) (.0124) (.0832)

[2.12] *WTP(h:no reminder)*=-43.91+15.786(*MARKET*)+.630(*SELL*)
(*F* statistic) (3.874) (4.670) (23.665)
(*p* value) (.0578) (.0383) (.0000)

[2.13] *WTP(a)*=-.096+4.363(*MARKET*)+.046(*SELL*)
(*F* statistic)(.0003)(4.670) (23.665)
(*p* value) (.9859)(.0177) (.2165)

The *F* critical value for all of these tests is 4.17. Thus, for both *WTP(h:reminder)* and *WTP(a)*, the constant is not significantly different from zero, *MARKET* is a significant determinant of WTP, and *SELL* is not a significant determinant of WTP at the 0.05 level. This contrasts with *WTP(h:no reminder)*, where the constant, *MARKET*, and *SELL* are all significant determinants of hypothetical WTP. Incidentally, the results for all three equations are consistent with the *t* statistics, which were reported in Loomis et al. (1996), although they are not valid because of non-normality.

Likelihood Ratio Tests of the Equality of Regression Coefficients for *WTP(a)*, *WTP(h:no reminder)*, and *WTP(h:reminder)* From Independent Treatments.

A likelihood ratio test of $H_0{:}B_0{=}C_0$; $H_0{:}B_1{=}C_1$; and $H_0{:}B_2{=}C_2$, where *Bs* are coefficients from equation [2.11] and *Cs* are coefficients from equation [2.13] yields a chi-square of 18.686.

A likelihood ratio test of $H_0{:}B_0{=}C_0$; $H_0{:}B_1{=}C_1$; and $H_0{:}B_2{=}C_2$, where *Bs* are coefficients from equation [2.12] and *Cs* are coefficients from equation [2.13] yields a chi-square of 76.5.

A likelihood ratio test of $H_0{:}B_{0[2.11]}{=}B_{0[2.12]}$; $H_0{:}B_{1[2.11]}{=}B_{1[2.12]}$; and $H_0{:}B_{2[2.11]}{=}B_{2[2.12]}$, where $Bs_{[2.11]}$ are coefficients from equation [2.11] and $Bs_{[2.12]}$ are coefficients from equation [2.12], yields a chi-square of 41.89.

The critical chi-square value for each of these tests is 7.82, indicating that the independent variables do not affect the dependent variables in the same way in the actual cash and hypothetical WTP treatments, regardless of whether the *reminder* is included or not.

Criterion Validity Tests With Paired Samples.

H_0:WTP_i(h:no reminder=WTP_i(a)).

As shown in the first row of Table 2.4, all three statistical tests reject the null hypothesis. In 28 of 35 cases, *WTP(h:no reminder)* exceeded *WTP(a)*, and there were 7 ties.

H_0:WTP_i(h:reminder=WTP_i(a)).

As shown in the second row of Table 2.4, all three statistical tests reject the null hypothesis. In 17 of 33 cases, *WTP(h: reminder)* exceeded *WTP(a)*. There were 14 ties.

Table 2.4.

Probability Levels Associated With Testing the Equality of Actual and Hypothetical WTP (With and Without Reminder) From Paired Samples

Hypothesis Regarding Paired Responses	Sign Test of the Medians	PTMP (Test of Means)	PTMP (Test of Medians)
WTP(h:no reminder)=WTA(a)	p=0.0000	p=0.0003	p=0.000002
WTP(h:reminder)=WTP(a)	p=0.0007	p=0.00125	p=0.00035

An OLS regression of equation [2.5] yielded:

[2.14] $WTP_i(a)=8.79+.067(WTP_i(h:no\ reminder))$.

The Wald test of $H_0:B_0=0$ yields a p value of 0.0000, indicating that the constant is significantly different from zero and a p value of 0.0000 for $H_0:B_1=1$, indicating that the slope coefficient is significantly different from one. The Wald test of $H_0:B_1=0$ yields a p value of 0.0130, indicating that the slope coefficient is significantly different from zero. Although B_1 does not equal 1 without the *reminder*, the coefficient is still significant: every dollar of hypothetical WTP translates to $.07 of actual WTP.

An OLS regression of equation [2.6] yielded:

[2.15] $WTP_i(a)=-1.385+0.5634(WTP_i(h:reminder))$.

The Wald test of $H_0:B_0=0$ yields a p value of 0.6872, indicating that the constant is not significantly different from zero. The Wald test of $H_0:B_1=1$ yields a p value of 0.0000 indicating the slope coefficient is significantly different than one.[4] However, the Wald test of $H_0:B_1=0$ yields a p value of 0.0000, indicating that the slope coefficient is significantly

[4] With small sample sizes such as ours the Wald statistic may be inflated, resulting in too high a level of significance since in small sample sizes the Wald statistic is not precisely distributed χ^2. See Kennedy (1992) for more details.

different from zero. We interpret this to mean that although B_1 does not equal 1 with the *reminder*, the coefficient is still significant and every dollar of *WTP(h)* translated into $.56 of actual WTP.

The *reminder* appears to be effective in narrowing the gap between hypothetical and actual WTP responses. Compare the coefficient on B_1 of 0.56 with the *reminder* to 0.067 in the hypothetical with *no reminder*, and the constant term with the *reminder* is not statistically different from zero. However, the overall results of the regression tests suggest rejection of H_0:$B_1=1$ for equations [2.14] and [2.15], indicating that hypothetical and actual WTP are not equal in the paired responses, regardless of whether the *reminder* was included or not.

In terms of correlations, the paired hypothetical and actual cash responses from treatment No. 2 have a 0.733 correlation which is significant at the 0.001 level, whereas the responses from treatment No. 1 have a correlation of 0.397, which is significant at the 0.05 level. Using a test of the difference between two correlations (Blalock 1972:309-10), these correlations are statistically different at the 0.05 level, indicating that use of the *reminder* significantly improved the correlation between actual and hypothetical WTP.

DISCUSSION AND CONCLUSION

Our results are consistent with those of Neill et al. and Kealy et al. in finding that hypothetical WTP was significantly greater than actual cash WTP both with and without the *reminder*. Excluding WTPs over $1,000, the Neill et al. map experiment produced differences between hypothetical and actual WTP that were about 9:1; however, their painting experiment produced between hypothetical open-ended CVM WTP and actual cash

second-price Vickrey auction WTP of 3.89. Our most comparable treatment (*no reminder*), which used nearly the exact wording of Neill et al., produced differences in WTP that were about 3:1. In our hypothetical session with the *reminder*, the difference between the hypothetical and real WTP is 1.8:1. With the *reminder*, the highest hypothetical WTP was $100.

The reduced discrepancy of actual and hypothetical WTP between our experiment and Neill et al.'s may be due in part to the wording of the *reminder* made after extensive pretesting and debriefing with pretest respondents in the paired samples which provided additional insight about how they determined their WTP in the hypothetical and actual markets. The disparity between hypothetical and actual WTP was reduced to 1.8:1 by reminding respondents: (a) not to bid what they thought they good might sell for in a store; (b) to act as if they were in a real market; and (c) to take their household budget and available funds into consideration when bidding, although the reduction was not large enough to be statistically significant. There are more fundamental reasons for the differences between hypothetical and actual cash WTP than those addresses by our *reminder*. The *reminder* nevertheless resulted in a significant improvement in the correlation of actual to hypothetical WTP, and the regression analyses showed that the *reminder* lessened the association of hypothetical WTP with perceived market price.

One reason why the overall effect of the *reminder* may not have been significant is that in retrospect there seems to have been some ambiguity in the instructions to the respondents. Specifically, on the one hand we ask them not to tell us what they thought a fair market price would be or what they thought "The Shadow Hunter" might sell for in a store,

and on the other hand we asked them to act as if they were in a real market. Well, if one were in a real market, one would expect to pay a fair market price. This may have caused some cognitive dissonance in any respondents who also perceived these instructions as inconsistent and ambiguous and negated the effect of the *reminder* by causing individuals who were not in the *MARKET* for art prints to behave as they were in the *MARKET* for art prints when in fact they were not. The statement that instructed the respondents to behave as if they were participating in a real sealed-bid auction came after the actual "reminder" statement and attempted to reiterate the "reminder" statement to the respondents. Eliminating the reiteration from the *WTP(h:reminder)* questionnaire used in Treatment No. 2 would eliminate this ambiguity.

With *no reminder*, the highest bids were $150 and $400, and inclusion of these two high bids explains much of the difference between our two hypothetical treatments. Thus, one advantage of the *reminder* is that it apparently reduces the tendency of some respondents to give unrealistically large WTP bids.

Since we found that perceived market price of the print was significantly related to hypothetical WTP with *no reminder*, it may be that some subjects attempted to rely on their estimate of what they thought the price of the good was to estimate their WTP in the hypothetical treatment. This finding raises the possibility that experiments using market goods may not provide an unambiguous test of criterion validity of CVM for non-market goods, as has been previously assumed.

CHAPTER THREE

EVALUATING THE VALIDITY OF THE DICHOTOMOUS CHOICE QUESTION FORMAT IN CONTINGENT VALUATION[8]

[8]This chapter draws largely upon: Loomis, John, Thomas Brown, Beatrice Lucero, and George Peterson. 1997. "Evaluating the Validity of the Dichotomous Choice Question Format in Contingent Valuation." *Environmental and Resource Economics* 10:109-23; however, it departs significantly in some respects and provides more information on the experiment and data analyses.

INTRODUCTION

In this chapter, WTP from the hypothetical and actual cash DC treatments are compared with each other and with WTP from hypothetical and actual open-ended treatments to address the following objectives: (1) test whether there is a significant difference between hypothetical and actual cash WTP for a private good with the DC question format; (2) compare the performance of the open-ended and DC WTP question formats in hypothetical and actual cash treatments for the same private good; and (3) test for starting point bias in the DC question format.

RESEARCH DESIGN

Treatments.

The results reported in this chapter involve four independent treatments:

No. 2: $WTP^{OE}(h:reminder)$: hypothetical open-ended WTP with "reminder;"

No. 3: $WTP^{OE}(a)$: actual cash open-ended WTP;

No. 4: $WTP^{DC}(h:reminder)$: hypothetical DC WTP with "reminder;" and

No. 5: $WTP^{DC}(a)$: actual cash DC WTP.

"Treatment No. 1" is not used here, as it continues to be identified as Treatment No. 1 in Chapter One. Treatments No. 2 and 3 are the same treatments No. 2 and 3 discussed in Chapter One, respectively. Treatments No. 4 and 5 were introduced in Table 1.1.

Null Hypotheses.

The null hypotheses to be tested are:

[3.1] $H_0: WTP^{DC}(h:reminder) = WTP^{DC}(a);$

[3.2] $H_0: WTP^{DC}(a) = WTP^{OE}(a);$

[3.3] $H_0: WTP^{DC}(h:reminder) = WTP^{OE}(h:reminder);$ and

H_0:B_1=0 in equations [3.4] and [3.5]:

[3.4] $WTP_i^{OE}(h{:}reminder)=B_0+B_1(\$BID)+B_2(MARKET)+B_3(SELL)$;

[3.5] $WTP_i^{OE}(a)=B_0+B_1(\$BID)+B_2(MARKET)+B_3(SELL)$

where $WTP_i^{OE}(h{:}reminder)$ and $WTP_i^{OE}(a)$ are both follow-up responses from the open-ended treatments with which we immediately followed the hypothetical DC treatment. These last two null hypotheses test for starting point bias.

EXPERIMENTAL DESIGN

Treatments No. 4 and 5 were applied as a part of the overall experiment discussed in Chapter One and reflected in Table 1.1. The manner in which these two treatments were executed is the same as that described in detail in Chapter Two for treatments No. 1, 2, and 3. The sample sizes for treatments No. 2 and 3 were 33 and 32, respectively. Because DC CVM only records whether an individual's WTP is greater or less than their bid amount, the DC CVM treatments required larger samples. Although we started out with sample sizes of 56 in both treatments No. 4 and 5, which we divided into two smaller groups of 28 each, we somehow lost two observations in treatment No. 4. This sometimes happens when respondents do not provide a response to all questions. All of the sessions were conducted by the same researchers, who closely followed a script.

The participants in treatments No. 4 and 5 were assigned from the same larger pool of participants from which respondents were randomly assigned to treatments No. 1, 2, and 3. The same good was used in all five treatments – a signed wildlife art print of a wolf standing in a forest called *"The Shadow Hunter"*. The wording of the questionnaires used in treatments No. 2 and 3 was presented in full in Chapter Two and will not be reproduced here. The wording of the questionnaires for the newly introduced treatments is as follows.

Wording of WTP Questionnaires.

Treatment No. 4. The wording of the *WTP^{DC}(h:reminder)* questionnaire was:

3. You are being asked to participate in a ***HYPOTHETICAL*** *sealed bid auction for this art print. We would like to know if you would pay the dollar amount in question #4 below to take this art print with you at the end of this session, if this one art print were actually for sale.*

At this time in the survey, we are NOT asking what you think the art print might sell for in a store or what you think its fair price is. Rather, we want to know whether you would honestly be prepared to pay the dollar amount stated in question #4 below right now to buy the art print you are being shown, if you would really be required to pay your bid amount with cash, write a check today, or sign a Promissory Note payable on or before August 19. Please take into consideration your budget and what you can afford to pay. If the price in question #4 is different from what you judge a fair price to be, that is OK. We want to know if you would actually be prepared to pay the price listed in question #4 for the art print.

Take a few moments to think about whether you honestly would be prepared to pay the printed dollar amount for this art print if it were being offered for sale to you today. Although the question is hypothetical, we want you to answer as if it were for real - as if you were participating in a real sealed-bid auction and would really be required to pay the printed dollar amount.

STOP. Do not answer questions #4 and #5 until instructed to do so.

If only one person answers YES, he or she would have obtained the print at the stated price on the survey. If there is more than one person stating YES we will have additional questions to determine who would have been the highest bidder.

4. *Would you really be prepared to pay $_____for this art print?*

   _____ *YES, I would pay this amount.*

   _____ *NO, I would not pay this amount.*

5. *Please describe what factors you considered when arriving at your dollar amount.*

Treatment No. 5. The wording of the $WTP^{DC}(a)$ questionnaire was:

*3. As a part of this experiment, we are now going to conduct a **REAL** auction. If you wish to actually buy the art print at the price stated below, answer YES in question #4. If you are the only person who answers YES, you will be required to buy the art print at the stated price. If there is more than one person stating YES, we will have additional questions to determine the highest bidder.*

We will accept cash or check (payable to C.S.U.) for your purchase. We understand that you may not have anticipated the need to bring cash or your checkbook with you today, so we will also accept a signed Promissory Note payable on or before August 19.

In any case, the successful buyer will be able to take the art print with them at the end of this session.

Now take a few moments to think about what having this art print would be worth to you and then answer questions #4, #5, and #6.

If you want to buy the art print at the stated price on the sheet, answer YES. If you don't want to purchase the art print at this price, answer NO.

4. Are you prepared to pay $_____for this art print?

_____ *YES, I will pay this amount.*

_____ *NO, I will not pay this amount.*

5. If I am the highest bidder I promise to complete the purchase at this bid amount.

_____
Signature

6. *Please describe what factors you considered when arriving at your decision.*

_____

In both the hypothetical and actual cash DC treatments, each person's answer sheet contained one of ten different prices ranging from $2 to $120, but centered around the mean of the pre-test open-ended WTP responses, $38.

DISCUSSION OF STATISTICAL TESTS

Several approaches are available for testing the equality of hypothetical and actual cash WTP estimated from DC questions administered to two independent samples. The most direct test of the raw data is to test the differences in percentage of NO responses to each bid level using the Mielke and Berry (1985) non-asymptotic approach based on the chi-square statistic for equality of cell frequencies. This test is used rather than a chi-square due to the very low cell frequencies at each bid amount occurring from our small sample sizes in each treatment. Specifically, some of the cells in our contingency table have fewer than five frequencies, which invalidates use of the asymptotic chi-square statistic. The non-asymptotic approach developed by Mielke and Berry takes into consideration the exact, mean, variance, and skewness of the distributions in analyzing data with low and/or highly uneven marginal frequencies in comparing the percentage of NO responses at each bid amount between the actual and hypothetical treatments. We also test differences in the percentage of YES responses in the same manner.

Loomis et al. (1997) report another approach for testing the raw data to determine whether the odds of agreeing to pay the bid amount (*[(Pr(Yes)/(1-Pr(Yes)]*) are equal between the hypothetical and actual DC treatments using the Cochran-Mantel-Haenszel nonparametric contingency table test and reference Snedecor and Cochran (1980:210-13).

> The Cochran-Mantel-Haenszel test is an archaic method for combining independent one-sided P-values associated with a number of tests such as binomial and 2 by 2 contingency table tests. This test requires the unnecessary and erroneous assumption that the standardized deviates follow a normal distribution (the tests in question are discrete, i.e., not continuous as presumed). For this purpose, the exact one-sided P-values (assuming discrete tests such as the binomial or 2 by 2 contingency table tests are used) based on a Fisher's exact test may be combined using Fisher's method for combining P-values (also and exact tests requiring no unnecessary assumptions. (Mielke 2003; Mielke and Berry 2001).

The Cochran-Mantel-Haenszel test basically employs a "correction for continuity" in order to widen both the variance and confidence intervals of the normal distribution. (Snedecor and Cochran: 210-13; 253-56).

Maximum WTP is not directly observed in the DC approach but it can be estimated parametrically or calculated nonparametrically. Parametrically, the two most common approaches are Hanemann's (1984) utility difference and Cameron's (1988) variation function. McConnell (1990) has shown that these two approaches are equivalent with linear specifications of the random utility model and constant marginal utility of income.

Hanemann (1984) views CVM respondents using a utility difference approach when they decide whether to answer "YES" or "NO" at the stated bid amount *($BID)*, i.e., the difference in their level of utility if they were to answer "YES" and if they were to answer "NO". Hanemann introduces a random error term in each of these utility levels; however, these error terms are assumed to have a mean of zero, and the difference in these error terms is assumed to be symmetric. If the cumulative distribution function of this difference in error terms is logistically distributed, a logit model of the probability of a YES response is related to the respondent's bid amount *($BID)* and attitude/demographic variables *(Z)* as in equation [3.6]:

$$[3.6] \quad \log\left[\Pr ob(YES)/(1-\Pr ob(YES))\right] = B_0 + B_1(\$BID) + B_2(Z_1 + \ldots + B_n(Z_n).$$

WTP is the area under the cumulative distribution function (CDF or g(BID)), as follows:

$$[3.7] \quad WTP = \int_0^{\infty}\left[1-g(\$BID)\right]\partial(\$BID) - \int_{-\infty}^{0} g(\$BID)\partial(\$BID).$$

When WTP is restricted to positive values, this formula reduces to:

[3.8] $$WTP = \int_0^{\infty} [1 - g(\$BID)]\partial(\$BID).$$

To calculate the mean WTP from the untruncated[9] logistic distribution, the formula for the mean of a non-negative[10] random variable is used (Hanemann 1989):

[3.9] $$MeanWTP = 1/B_1^{*}(\ln(1 + \exp(B_0 + \sum(B_n(Z_n))))).$$

The median is provided by:

[3.10] $$MedianWTP = (B_0 + \sum(B_n(Z_n)))/B_1$$

where B_n is the vector of coefficients and Z_n are the sample means of the associated independent variables. Although Loomis et al. (1997) report equations [3.9] and [3.10] as Hanemann's respective formulas for mean and median WTP when WTP>0, Hanemann (1984; 1989) reports equation [3.9] as the mean and median for WTP when WTP is restricted to positive values and equation [3.10] as the mean and median for WTP when WTP is not restricted to positive values. In the Hanemann "no income effects" utility model, the mean and median coincide, which is to be expected under assumptions of symmetric error terms with means of zero. This assumption of symmetry further implies that *g($BID)* in [3.7] and [3.8] are symmetric. (1984.) In the Hanemann "no income effects" linear random utility model, the median is determined by restricting the probability of a "YES" response to .50, which implies that the respondents were just on the verge of responding NO (YES) but instead responded YES (NO) at whatever *$BID* amount they received randomly. (1984.)

---

[9] Loomis et al. (1997) incorrectly report that mean WTP was estimated from the "truncated" logistic distribution. We did not truncate the CDF at the highest bid amount of $200.

[10] Although this is reported to be the formula for the median of a "non-negative random" variable, it actually restricts the variable to positive values, i.e., the Hanemann estimates include no zero bids, but the raw open-ended data for treatments No. 1, 2, and 3 include many zero bids.

Equation [3.10] does not restrict WTP to positive values. Indeed in treatments No. 3 and 4, equation [3.10] generated 9 and 15 negative values of median WTP as high as ($34.10) and ($14.54), respectively.

If the cumulative distribution function of the difference in error terms is assumed to be normally distributed, then the probit model is estimated. The probit model is given by:

$$[3.11] \quad F^{-1}(\pi_i) = B_0 - B_1(\$BID) + B_2(Z_1) + \ldots + B_n(Z_n)$$

where $F^{-1}\pi_i$ is the inverse of the standard normal cumulative distribution function (Kmenta 1986:553). Median WTP is calculated as in equation [3.10].

Mean WTP can also be calculated using a nonparametric approach proposed by Kristrom (1990). Based on the proportion of YES responses at each bid amount, the area under what Kristrom calls the "empirical survival function" is calculated.

Another approach involves a comparison of confidence intervals about estimates of $WTP^{DC}(a)$ and $WTP^{DC}(h{:}reminder)$. This was done for the means estimated parametrically based on the approach of Park et al. (1991) and nonparametrically using the approach of Duffield and Patterson (1991). If the confidence intervals do not overlap, we may conclude that hypothetical and actual WTP are different.

A test of coefficient equality of the logit equations used to estimate WTP was also performed. As shown in equation [3.9], WTP depends directly on the estimated coefficients in the logit equation. We test for equality of the logit coefficients using a log likelihood ratio (LLR) test. The null hypothesis in our LLR test is: $B_n(h) = B_n(a)$, where $B_n(h)$ and $B_n(a)$ are the coefficients in the logit equations for hypothetical WTP and actual cash WTP, respectively. The test is carried out by comparing the log likelihood of a single logit equation

(restricted model) estimated by combining or pooling observations from both the actual cash and hypothetical WTP responses to the sum of the log likelihood of the two (cash and hypothetical) logit equations estimated separately (unrestricted model). The LLR test follows a chi-square distribution with degrees of freedom equal to the sum of the number of coefficients in the two unrestricted models minus the number of coefficients in the restricted model.

Another approach involves the method of convolutions, which employs a formal statistical test of the differences in empirical WTP distributions derived from DC data (Poe et al. 1994). These authors note that their method is less prone to type II error than a comparison of confidence intervals and more relevant to comparisons of mean WTP than the LLR test of logit coefficient equality. The method of convolutions determines if there is a statistically significant difference between two simulated WTP distributions. According to Poe et al. (907), it accommodates any distributional form. The method involves calculating the probability of all possible differences (convolutions) between discrete values in the two distributions. The method then tests whether the $(1-\alpha)$confidence interval for this convolution or set of differences includes zero. In addition, the method calculates an alpha level for rejecting the null hypothesis of equality of the two distributions.

Loomis et al. (1997) also report Mann-Whitney tests of the median when comparing open-ended and discrete choice data, although these tests were also not part of the initial data analyses because they are not commonly used to test discrete data. The Mann-Whitney test was considered inappropriate for testing our open-ended data for all of the reasons discussed

in detail in Chapter Two. The appropriateness of this test for estimated open-ended and DC WTP data depends upon the normality and continuity of the estimated data.

Finally, the test of Boyle, Bishop, and Welsh (1985) is used to test for starting point bias. In particular, starting point bias is present in the sample if B_1 is significant in the regression of equation [3.4] or [3.5].

RESULTS OF HYPOTHESES TESTS AND ESTIMATED EQUATIONS

Comparison of Demographics Across Sessions.

Before comparing estimates of WTP, we tested whether the four samples were significantly different from each other in terms of standard demographics. One-way ANOVA's were performed for education (F=1.81, p=0.15), age (F=0.92, p=0.43) and income (F=1.28, p=0.282). As indicated by the p values, the samples are not significantly different.

Direct Tests of Raw Data.

H_0: WTP^{DC}(h:reminder)=WTP^{DC}(a).

The frequency of NO responses for these two treatments are shown in the (2 x 10) contingency table reflected in Table 3.1.

TABLE 3.1

CONTINGENCY TABLE OF NO RESPONSES

Bid	$2	$4	$8	$12	$18	$24	$36	$48	$72	$120
NO in Hypothetical	0	2	2	1	4	4	3	3	4	6
NO in Actual	1	2	3	4	3	5	5	6	6	6

The most direct, powerful, and appropriate test of the raw data is Mielke and Berry's non-asymptotic approach based on the chi-square statistic for equality of cell frequencies. Applying this test to the NO responses yields a p value of 0.9753, indicating no significant

difference in the percentage of NO responses between $WTP^{DC}(h:reminder)$ and $WTP^{DC}(a)$. Applying the same test for the YES responses reflected in Table 3.2 yields a p value of 0.7719, also indicating no significant difference in the percentage of YES responses between $WTP^{DC}(h:reminder)$ and $WTP^{DC}(a)$. The contingency table for this calculation necessarily excludes the \$72 and \$120 categories, as no one responded YES to these dollar amounts.

TABLE 3.2

CONTINGENCY TABLE OF YES RESPONSES

Bid	\$2	\$4	\$8	\$12	\$18	\$24	\$36	\$48	\$72	\$120
YES in Hypothetical	5	3	4	4	2	2	3	2	0	0
YES in Actual	5	4	2	2	1	0	1	0	0	0

Applying Fisher's exact test to the same contingency tables confirms these results. Specifically, using the (2 x 10) contingency table for NO responses, the p value is 0.9801; and using the (2 x 8) contingency table for YES responses, the p value is 0.8493.

The Cochran-Mantel-Haenszel contingency table test, however, rejects the null hypothesis. The p value of 0.018 indicates that the probability of a respondent providing a YES response is significantly greater in the hypothetical treatment than in the actual cash treatment.

Hypotheses Tests Based Upon Estimated Equations.

In order to calculate WTP and perform the statistical comparisons of hypothetical and actual cash behavior, logit equations must be estimated for the two DC treatments because WTP is not directly observed with the DC question format. The following tests are all based upon WTP estimated according to the approach of Hanemann (1984, 1989).

H_0: WTP^{DC}(h:reminder)=WTP^{DC}(a).

In the initial data analyses, the logit equations for treatments No. 3 and 4 were estimated with specifications analogous to the open-ended equations [2.11] and [2.13] presented in Chapter Two, respectively, as follows:

[3.12] *YPAY(hyp)=-4.5560-.0938($BID)+1.7904(MARKET)+.0185(SELL)*; and
(*t* statistic) (-2.51)(-2.757) (2.95) (1.44)

[3.13] *YPAY(a)=-2.0983-.1269($BID)+1.1370(MARKET)-.0022(SELL)*,
(*t* statistic)(1.52)(-2.59) (2.40) (-.50)

where *YPAY* is the log odds ratio. The *t* statistics are reported here because the WTP distributions for both of these equations, estimated as described above, test normal by the Jarque-Bera normality test. The *t* tests for these logit regressions indicate that in both the hypothetical and actual DC treatments, *MARKET* is a significant determinant of the respondents' respective decisions whether or not to pay the randomly distributed *$BID* amounts ranging from $2-$120 and centered around a mean of $38; and the amount they thought *"The Shadow Hunter"* would *SELL* for in a store is not a significant determinant. A LLR test of the equality of these coefficients yields a chi-square of 10.32. The critical chi-square for a 0.05 level of significance is 9.448 and for a 0.01 level is 13.277. Thus we reject the null hypothesis that the independent variables affect the dependent variable of equation [3.12] in the same manner that they affect the dependent variable of equation [3.13] at a 0.05 level of significance, but we cannot reject equality at the 0.01 level.

Subsequent to this initial analyses, equations [3.14] and [3.15] were estimated for WTP^{DC}*(h:reminder)* and WTP^{DC}*(a)*, respectively, with alternative specifications to those of equations [3.12] and [3.13].

[3.14] *YPAY(hyp) = -10.77-0.2578($BID)+1.96(MARKET)+7.84(LIKE)-0.537(AGE)+0.09(INC)*
(*t* statistic) (-1.75) (-2.38) (1.85) (2.27) (-2.12) (1.55)

This logit equation's goodness of fit statistic, the chi-square, equals 56.6 which is significant at the 0.01 level.

[3.15] *YPAY(a) = -7.92-0.1787($BID) + 1.44(MARKET)+1.37(LIKE)-0.04(AGE)+0.05(INC)*
(*t* statistic) (2.05) (-2.56) (2.47) (1.88) (-0.88) (1.36)

This logit equation's chi-square equals 36.6 which is significant at the 0.01 level. A LLR test of the equality of the logit coefficients in equations [3.14] and [3.15] yields a chi-square of 16.11, to be compared with critical chi-square values of 12.59 at the 0.05 significance level and 16.812 at the 0.01 level. Thus we reject the null hypothesis that the independent variables affect the dependent variable of equation [3.14] in the same manner that they affect the dependent variable of equation [3.15] at the 0.05 significance level, but we cannot reject equality at the 0.01 level.

Equations [3.16] and [3.17] provide the probit equations for the hypothetical and actual DC treatments, respectively.

[3.16] *YPAY(hyp) = -6.11-0.1499($BID)+1.15(MARKET)+4.51(LIKE)-0.31(AGE)+0.05(INC)*
(*t* statistics) (-1.75) (-2.42) (1.83) (2.28) (-2.10) (1.49)

[3.17] *YPAY(a) = -4.11-0.0809($BID)+0.745(MARKET)+0.678(LIKE)-0.02(AGE)+0.02(INC)*
(*t* statistics) (2.05) (-3.12) (2.57) (1.82) (-0.68) (1.16)

In both the logit and probit models, the *$BID* variable is significant and negatively related to the log odds ratio in both hypothetical and actual markets, whereas being in the *MARKET* and liking (*LIKE*) the good increased the log odds ratio. The pattern of variable significance is identical between the logit and probit models, however, the coefficients cannot be interpreted the same as one interprets OLS coefficients or compared to each other in the same manner.

Thus, the LLR tests of equality of the logit coefficients for all three model specifications support rejecting the null hypothesis at the 0.05 level of significance but not at the 0.01 level. The method of convolutions for comparing WTP distributions estimated from hypothetical and real markets also indicates we should reject equality at the 0.05 level but not at the 0.01 level. As can be seen in Table 3.3, the 95% confidence intervals (CI's) for hypothetical and actual WTP derived from the logit estimates of the DC responses overlap in the tails. The Mann-Whitney test, however, yields p=0.006, and therefore rejects equality.

Table 3.3

Descriptive Statistics:
Comparison of Hypothetical and Actual WTP

Treat-ment	Estimator	Sample size	Mean (Median) WTP by treatment [95% CI]			
			Dichotomous Choice		Open-ended	
			Hypothetical	Real Mkt.	Hypothetical	Real Mkt.
No. 2	OLS	33			26(20) [19-33]	
	Nonparametric				26(20) [17-35]	
No. 3	OLS	32				14(14) [12-16]
	Nonparametric					14(10) [10-18]
No. 4	Logit	52	31(28) [20-37]			
	Nonparametric		33 [20-46]			
No. 5	Logit	55		12(9) [6-22]		
	Nonparametric			11 [9-14]		

Table 3.3 summarizes the means and 95% confidence intervals (CI) for all four treatments. The parametric results represent estimates calculated from the logit and OLS

regressions for DC and open-ended responses (subsequently explained), respectively, with specifications equivalent to those of equations [3.16] and [3.17]. The nonparametric results represent Kristrom's estimator for the DC and the average of the raw data for the open-ended responses. Confidence intervals were calculated parametrically according to the approach of Park et al. and nonparametrically using the approach of Duffield and Patterson. Estimates of WTP from the probit models are not shown, as they are not statistically different from the logit estimates. "[T]he logistic and probit formulations are quite comparable, the chief difference being that the logistic has slightly flatter tails..." (Gujarati. 1995:559.) Indeed, all of the distributions for the logit estimates based on Hanemann's approach test normal by the Jarque-Bera normality test, although the raw open-ended data for treatments No. 1, 2, and 3 do not.

Although the WTP estimates generated by equation [3.10] are reported as medians in Table 3.3, as they were in Loomis et al. (1997), they are actually equivalent means and medians when WTP is not restricted to positive values, as previously explained. The difference between the WTP estimates generated by equations [3.9] and [3.10] is most pronounced in the lower end of WTP estimates. Equation [3.10] generated 9 negative values in treatment No. 3 and 15 in treatment No. 4. Initially in the positive spectrum of the estimates, the values generated by equation [3.10] are lower than those generated by equation [3.9], but by about the middle of the range, the estimates become identical or near identical. This is why the reported "median" values are consistently lower than the mean values.

$H_0: WTP^{DC}(a) = WTP^{OE}(a)$.

To test whether is a significant difference between open-ended and dichotomous choice values in hypothetical and actual cash treatments we must use estimated WTP from

the OLS and logit equations. This is because WTP cannot be directly observed in the dichotomous choice case, so equivalent treatment requires use of estimated WTP from fitted models in both cases. For the hypothetical and actual dichotomous choice treatments, the logit equations given in [3.16] and [3.17], respectively, were used. To calculate each individual's WTP from these equations, we inserted individual *i*'s values for *MARKET*, *LIKE*, *AGE*, and *INC* into equation [3.6] and added this to the constant term B_0, to arrive at a grand constant that is used in equation [3.9]. Equation [3.18] represents a WTP equation estimated from an OLS equation based on the data from treatment No. 4 and using a variable specification equivalent to that of logit equation [3.17]:

[3.18] $WTP^{OE}[a(EST)] = -3.09 + 3.71(MARKET) + 1.36(LIKE) - 0.022(AGE) + 0.064(INC).$
(*t* statistics) (-0.18) (1.41) (0.36) (-0.08) (0.82)
R^2=0.21 F=1.76

As indicated by the very low R^2, this estimated equation is very poorly fit. None of the explanatory variables is significant - not even MARKET, which was a significant determinant of WTP in treatments No. 1, 2, and 3 when the raw data was tested.

The Mann-Whitney test of medians yields a *p* value of 0.77, also indicating that we cannot reject equality of $WTP^{DC}(a)$ and $WTP^{OE}(a)$. The CI for $WTP^{OE}(a)$ from either the parametric (mean of the estimated data) or nonparametric (mean of the raw data) are completely contained within the CI for $WTP^{DC}(a)$ using the parametric approach, suggesting that actual WTP estimated using the two question formats are not significantly different. Consistent with Kealy et al., our actual cash open-ended WTP was greater than our actual cash DC WTP, although in our case the difference is not significant at the 0.05 level.

H_0: WTP^{DC}(h:reminder)=WTP^{OE}(h:reminder).

Equation [3.19] represents a WTP equation estimated from an OLS equation used to estimate WTP based on the data from Treatment 3 and using a variable specification equivalent to that of logit equation [3.16]:

[3.19] WTP^{OE}[h:reminder(EST)] = -72.1 + 3.32(MARKET) +12.9(LIKE) +1.1(AGE) -0.094(INC)
(*t* statistics) (3.58) (1.04) (2.87) (2.90) (-0.67)
R^2=0.59 F=9.93

Although the R^2 indicates that this equation is better fit than [3.18], it is still not very well behaved. Not only is *MARKET* not a significant variable, but the coefficient for *INCOME* is negative, although it is not significant.

The Mann-Whitney test of medians yields a *p* value of 0.48 indicating that we cannot reject the null hypothesis. The CIs for WTP^{DC}(h:reminder) are nearly identical to the CIs for WTP^{OE}(h:reminder). Thus the null hypothesis cannot be rejected according to this approach.

The results indicate that the parametric logit estimate of WTP^{DC}(h:reminder) based on Hanemann's approach is 2.5 times WTP^{DC}(a) and 2 times WTP^{OE}[a(EST)]. The nonparametric estimate of WTP^{DC}(h:reminder) is 3 times the estimate of WTP^{DC}(a) and 2.3 times WTP^{OE}[a(EST)]. A similar relationship is evident from examining the differences in median WTP between treatments.

H_0: B_1=0 in equations [3.4] and [3.5].

Finally, our test of starting point bias in the follow-up open-ended WTP questions to the initial DC questions is evaluated for both hypothetical and actual cash WTP. Our regression of equation [3.4] yielded:

[3.20] WTP_i^{OE}(h:reminder)=-28.88+.06($BID)+10.74(MARKET)+.24(SELL)
(*p* value) (.0049)(.5146) (.0001) (.0001)

where the numbers in parentheses are p values associated with a chi-square when the significance of the constant and coefficients was tested by the Wald test. Both MARKET and SELL are significant determinants of hypothetical WTP in the DC treatment; however, $BID is not, so we conclude that there is no starting point bias in the hypothetical data.

Regression of equation [3.5] yielded:

[3.21] $WTP_i^{OE}(a)=-11.24-.02(\$BID)+6.12(MARKET)+.07(SELL).$
(p value) (.7512)(.2789) (.0207) (.9243)

In the actual cash open-ended follow-up to the dichotomous choice treatment, *MARKET* is a significant determinant of WTP, but *SELL* is not. *$BID* is also not a significant determinant of WTP in the follow-up treatment, so we also conclude that there is no starting point bias in the actual cash data. These findings of no starting-point bias using the DC question format were not presented in Loomis et al. (1997).

DISCUSSION AND CONCLUSIONS

The two valid direct tests of the raw data, Mielke and Berry's non-asymptotic approach based on the chi-square statistic for equality of cell frequencies and Fisher's exact test, both indicate no significant difference between hypothetical and actual cash WTP with the DC question format. These tests had been performed as a part of the initial data analyses, however, they were not reported in Loomis et al. (1997). The Cochran-Mantel-Haenszel test was not a part of the initial data analyses and is not a valid test because of its reliance upon an assumption of normality, as discussed in detail in a previous section. The valid and significant findings of equality are further supported by all of the other findings based upon the estimated equations, except for the Mann-Whitney test of medians. Specifically, the LLR test of equality of the logit coefficients for all three model specifications and the method of

convolutions all indicate that we can reject equality at the 0.05 level of significance but not at the 0.01 level, and the CI approach indicates that the estimates of actual and hypothetical WTP overlap in the tails. Thus, there is a range where actual and hypothetical WTP are equal, and Mielke and Berry's non-asymptotic chi-square test and Fisher's exact test are both powerful enough to detect this equality in the raw data. Therefore, we cannot reject equality of the hypothetical and actual WTP measures when the DC question format is used. These findings are different from those of Kealy et al. (1993) and Brown et al. (1996) who found that hypothetical WTP was statistically greater than actual cash WTP regardless of whether the open-ended or DC format is used.

The null hypothesis H_0:$WTP^{DC}(h:reminder) = WTP^{DC}(a)$ was the only hypothesis investigated in this chapter that permitted testing of the raw data. All the other hypotheses require use of logit estimates based upon the approach of Hanemann (1984, 1989). The problem that this presents is that the Hanemann methodology imposes normality on the estimated data. Indeed, all the estimates generated by this approach are all normally distributed, which is certainly not the case in the raw data for any of the open-ended or DC treatments No. 1-5. Therefore, although the Mann-Whitney test of the medians is certainly not appropriate for testing the raw open-ended or DC data for all the reasons discussed in Chapter Two, it would seem to be appropriate for testing the estimated data because the estimation methodology generates normally distributed and continuous data with no zero values and very few tied values. Therefore, to the extent that we can accept the appropriateness of an estimation methodology that generates normally distributed continuous data from otherwise non-normally distributed and non-continuous data, we can accept the appropriateness of the Mann-Whitney test of medians for this estimated data.

The weakness of equations [3.18] and [3.19] seems to indicate that estimated WTP is not a very good estimate of the raw WTP estimates elicited with the open-ended question format. In addition, a Mann-Whitney test of the equality of median WTP based upon the raw data for treatments No. 2 and 3 yielded a p value of 0.09; however, this same test using WTP data estimated from equations [3.18] and [3.19] yields a p value of .04. This casts a shadow of doubt on the validity of Hanemann's estimation methodology. Indeed, a test of the equality of mean and median WTP from the raw data and the estimated data would be very informative. Given that the rejection of this null hypothesis by the Mann-Whitney test of the medians is based on estimated data, which does not appear to be valid, it does not seem reasonable to rely upon this test over Mielke and Berry's non-asymptotic approach to the chi-square statistic and Fisher's exact test, which are based on the raw data.

Perhaps a more appropriate estimation methodology would be based upon the Least Absolute Deviation regression procedure also developed by Dr. Paul W. Mielke and his colleagues at Colorado State University. It performs a regression of the medians and does not impose normality on the data, not even asymptotically.

> Least absolute deviation (LAD) regression is an alternative to ordinary least squares (OLS) regression that has greater power for thick-tailed symmetric and asymmetric error distributions. LAD regression determines the conditional median (a conditional 50^{th} quantile) of a dependent variable given the independent variable(s) by minimizing sums of absolute deviations between observed and predicted values.... LAD regression can be used anywhere OLS regression would be used but is often more desirable because it is less sensitive to outlying data points and is more efficient for skewed error distributions as well as some symmetric error distributions. (Cade and Richards 2001:2.)

Mielke argues that tests based upon Euclidean space are more powerful than tests based upon non-Euclidean space, because data is collected in Euclidean space, which is metric. (Slauson et al. 1991) Hanemann also argues that median tests are much more robust than mean tests

because the mean tests are "very sensitive to slight changes in the shape of the distribution resulting from different estimation methods or outliers in the data" (1984:339).

For the remaining hypotheses that do not permit testing of the raw data, there is no inconsistency between the findings of the Mann-Whitney test and the other tests based upon the estimated data. They all find no significant difference between open-ended and DC WTP, whether elicited under hypothetical or actual cash scenarios.

Also significant is the fact that we found no starting point bias with the DC question format. Further significant is the fact that with the DC question format, the variable *SELL* was not a significant determinant of estimated WTP in hypothetical or actual cash treatments. Thus it would appear that this study strengthens the arguments for use of the DC over the open-ended question format, as recommended by the NOAA panel. This finding is different from that of Loomis et al. (1997).

This study can also be compared with those of Bishop et al. (1992) and Brown et al., as they also obtained both DC and open-ended estimates of WTP in both hypothetical and actual markets and used independent samples for each of the four experimental conditions. The principal difference among the studies was that Bishop et al.'s and ours valued a private good whereas Brown et al.'s valued a public good. Perhaps most notably, all three studies found that *actual* WTP was roughly the same regardless of whether the open-ended or DC format was used (Bishop et al. do not report a statistical test, but the other two studies found no significant difference). With respect to hypothetical WTP, the evidence is more mixed, with our study and Bishop et al. showing no difference, but other private good (e.g., recreation) studies and public good studies showing the DC estimate exceed the open-ended estimate (see Brown et al. for a list of the studies).

CHAPTER FOUR

PAIRED COMPARISON ESTIMATES OF WILLINGNESS TO ACCEPT VERSUS CONTINGENT VALUATION ESTIMATES OF WILLINGNESS TO PAY[11]

[11] This chapter draws largely upon Loomis, John, G. Peterson, P. Champ, T. Brown, and B. Lucero. 1998. "Paired Comparison Estimates of Willingness to Accept Versus Contingent Valuation Estimates of Willingness to Pay." *Journal of Economic Behavior and Organization.* 35:501-15; however, it departs significantly in many respects.

INTRODUCTION

This chapter reports the results of our second experiment where estimates of WTA for the same art print used in the first experiment are derived through PC. A lower bound on WTA is estimated by asking the respondents to make a binary choice between two alternative gains, for example, $50 or a specified private good. If the good is selected, then the lower bound for WTA for the private good is greater than $50. These PC-derived estimates of WTA are compared with $WTP^{dc}(h{:}reminder)$ and $WTP^{dc}(a)$ in order to address the following objectives: (1) test whether there is a significant difference between CVM-derived WTP and PC-derived WTA; (2) test whether the chooser reference point employed by PC avoids the loss aversion proposed by Kahnemann et al. (1990) because the individual is comparing two alternative gains; and (3) test whether differences between PC-derived WTA and CVM-derived WTP are smaller than the historical gap between CVM-derived measures of WTA and WTP.

The good of interest, a signed (but not limited edition) commercially available art print of a wolf standing in a forest called *"The Shadow Hunter,"* has many substitutes. For most people, expenditures on unframed art prints such as ours represent a small fraction of their household's income. Given multiple substitutes and a small fraction of income devoted to purchase of unframed art prints, Randall and Stoll (1980) and Hanemann (1991) suggest equality of WTA and WTP.

Treatments.

The test results reported herein involve three independent treatments:

No. 4: $WTP^{DC}(h{:}reminder)$;

No. 5: $WTP^{DC}(a)$; and

No. 6: $WTA^{PC}(h)$.

Treatments No. 1, 2, and 3 are not used here because they continue to be identified as treatments No. 1, 2, and 3, as introduced in Table 1.1 and discussed in Chapters Two and Three. Treatments No. 4 and 5 are treatments No. 4 and 5 introduced in Table 1.1 and discussed at length in Chapter Three. Treatment No. 6 is the only new treatment introduced here; it is the paired comparison treatment used to elicit WTA measures for *"The Shadow Hunter"*.

Null Hypotheses.

The null hypotheses tested in this chapter are:

[4.1] $H_0: WTP^{DC}(h:reminder) = WTA^{PC}(h)$; and

[4.2] $H_0: WTP^{DC}(a) = WTA^{PC}(h)$.

EXPERIMENTAL DESIGN

Participants.

College clerical and administrative staff in academic and nonacademic units were recruited and paid $20 for attending one of the 45 minute sessions held on campus. The sessions were conducted before work, at lunch and after work. Treatments No. 4 and 5 had sample sizes of 52 and 55, respectively. The PC experiment involved 103 individuals in 14 sessions (the sessions were smaller due to the limited number of laptop computers available).

Structure and Conduct of the Paired Comparison Sessions.

The PC experiments were run with 6-9 people per session. All sessions were led by the principal investigator who followed a written script. The PC sessions required the participants to choose between receiving one of 12 sums of money and one of four goods,

shown just two at a time. Choices between two different amounts of money were omitted as being trivial comparisons. The four goods, including *"The Shadow Hunter,"* an ATT cordless telephone, dinner and beverage for two at a local restaurant, and two tickets to one local college football game (seats on the 20 yard line), were described on a "product sheet" that was given to each participant. The actual wildlife art print and telephone were displayed in the room. For the restaurant dinner and football tickets, the actual menu and football tickets (along with upcoming football schedule) were mounted on poster boards, which were displayed in the room. The boards and goods were shown up close to the participants and remained on display during the session. Twelve different dollar amounts ranging from $4 to $295 were included in the PC. The investigator led the participants through the experiment and provided instructions on using the computers. The basic format of each session involved the PC exercise, followed by debriefing questions and finally socioeconomic questions. The entire experiment lasted about 40-50 minutes and was performed on the laptop computers.

Every individual that began the session completed the session, although they were told (in writing) they could leave at any time. Participants were careful to follow directions and did not discuss their choices with others during the experiment. Observation of participants suggested they put a great deal of thought into their choices. Comments after the session suggested they were stimulated by the experience.

Wording of the Questionnaires and Dollar Bid Amounts.

The exact wording of the questionnaires used to elicit $WTP^{DC}(h:reminder)$ and $WTP^{DC}(a)$ was presented in Chapter Three and is not repeated here. The exact wording of the introduction used to elicit $WTA^{PC}(h)$ 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.*

In both the CVM $WTP^{DC}(h:reminder)$ and $WTP^{DC}(a)$ treatments, each person's answer sheet contained one of ten different prices ranging from $2 to $120, but centered around the mean of the pretest open-ended WTP responses, $38. In the PC data, the distribution of bids was similar to those in the CVM $WTP^{DC}(h:reminder)$ and $WTP^{DC}(a)$ treatments, except the lowest amount was $4 and the highest was $295. The range of bids for the PC experiment was increased because the PC experiment also included treatments for public goods, which are not reported here.

MECHANICS OF THE METHOD OF PAIRED COMPARISON

PC can be applied to valuation by specifying a choice set that consists of carefully defined public or private goods (including one or more target goods of special interest) and sums of money. The presentation of a series of paired comparisons between a good and a sum of money is automated by means of an interactive computer program that presents pairs of goods (or a good and a sum of money) from the chooser reference point and requires the respondent to make a choice.[12] To control for order effects, the program presents the binary

[12] The program is written in Fortran but was compiled to run on nearly any DOS based machine with 640K of RAM and a hard disk.

choices in random order to each individual. The computer program automatically records for each respondent an ordered matrix of binary choices, the sequence in which the respondent sees the pairs, and the number of circular triads produced by the respondent's choices.

Loomis et al (1998:505) claim that:

> "[a] circular triad is the product of inconsistent choice. Preference for A over B, B over C, and C over A produces a circular triad (i.e., A>B>C>A). On face value a circular triad implies failure of the transitivity axiom of utility theory. As well understood by psychologists and probability theoreticians, however, the apparent intransitivity may be random or systematic. Random intransitivity may occur when the individual is indifferent between elements in a pair. Since PC does not allow for an 'indifference choice' we believe the transitivities are switching behavior on the indifferent choices and do not violate the transitivity axiom in expected outcome sense."

This statement circumvents the real issue at stake here. Any microeconomic text book will specify three fundamental axioms of rational choice, including completeness, transitivity, and continuity. These three axioms specifically state as follows (Nicholson 1992:74.):

> *Completeness:* If A and B are *any* two situations, the individual can always specify exactly one of the following three possibilities:
> 1. "A is preferred to B,"
> 2. "B is preferred to A," or
> 3. "A and B are equally attractive."
>
> Individuals are consequently assumed not to be paralyzed be indecision: They completely understand and can always make up their minds about the desirability of any two alternatives. The assumption also rules out the possibility that the individual can report both that A is preferred to B and that B is preferred to A.
>
> *Transitivity:* If an individual reports that "A is preferred to B" and that "B is preferred to C," then he or she must also report that "A is preferred to C." This assumption states that the individual's choices are internally consistent. Such an assumption can be subjected to empirical study, and in the Application for this chapter, we report on some of these studies. Generally, they conclude that a person's choices are indeed transitive, but that conclusion must be modified in cases where the individual may not fully understand the consequences of the choice he or she is making. Since, for the most part, we will assume choices are fully informed, the transitivity property seems an appropriate assumption to make about preferences.

Continuity: If an individual reports "A is preferred to B," then situations suitably "close to" A must also be preferred to B.

The real issue here is that the method of PC, practiced as described herein, itself violates the fundamental axiom of completeness by not offering the respondents a choice of indifference between the two goods presented, which is responsible for the circular triads. In the development stages of this experiment, I reviewed the method of PC for consistency with economic theory and pointed out the problems presented by the fact that PC violates the fundamental axiom of completeness. However, I was informed that the choice of indifference had been offered in past PC experiments but had proven "inefficient" because the respondents had made the choice of indifference too often. Thus it was decided to proceed with the experiment in violation of the completeness axiom of rational choice, which is clearly a necessary condition for the axioms of transitivity and continuity. The implications of this decision are reflected and discussed in the ESTIMATED EQUATIONS section of this chapter.

With four goods and ten sums of money, given that choices between sums of money were not asked, the respondent makes 46 choices. The individual chooses between the two goods or the good and the sum of money by pressing the right arrow key for the good on the right or the left arrow key for the good on the left. The right/left order between goods and money were randomly varied, so that a particular good or amount of money could appear on the right or left side of the screen.

DISCUSSION OF STATISTICAL TESTS

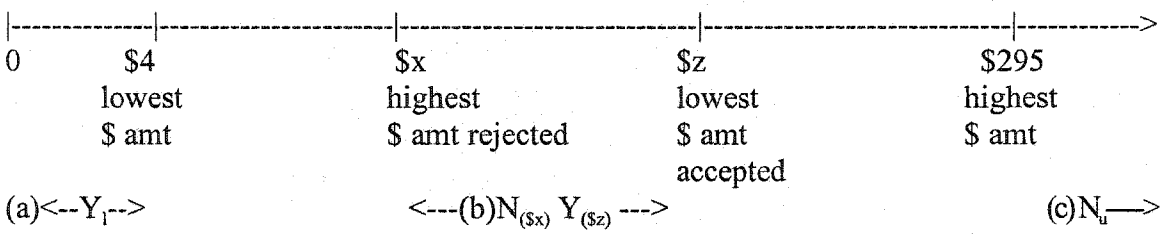
Estimating the Logit Equation and Calculating WTP and WTA.

Maximum WTP is not directly observed in the CVM DC approach, nor is minimum WTA directly observed in the PC method. In Chapter Three, Hanemann's (1984; 1989) utility difference approach to estimating maximum WTP with DC CVM. For the DC data, the logit model of the probability of a "YES" response, which is related to the amount the respondent is asked to pay *($BID)* and attitude/demographic variables *(Z)*, was presented in equation [3.6]; the cumulative distribution function for calculating WTP was presented in equation [3.8]; the formula for calculating the mean of a non-negative random variable from the truncated logistic distribution was presented in equation [3.9]; and the formula for determining the median WTP was presented in equation [3.10].

Calculation of WTA from PC data can be approached in at least two ways. First, one could analyze the data using a nonparametric approach. Since the PC method orders preferences among the goods and the dollars, WTA is between the lowest amount a respondent said he would accept and highest amount he would not accept for the good of interest. One could also construct an empirical cumulative distribution function from the raw responses and then use this to identify the median and calculate the mean.

The parametric approach to estimating minimum WTA from the PC data allows inclusion of covariates and explicitly incorporates deterministic and stochastic elements of the choice process. The approach used in this chapter is an adaptation of the multiple bounded method developed by Welsh and Bishop (1993). With this method each individual's responses are scanned to find the two dollar amounts where the individual switched from a *no* (N), would not choose that amount of money over the good, to a *yes* (Y),

would choose the money instead of the good. As shown below, there are essentially three possible outcomes: (a)Y_l; (b) $N_{(\$x)}$ $Y_{(\$z)}$; and (c) N_u. Category (a) arises when the individual chooses the lowest amount of money offered ($4 in our experiment) over the art print; category (b) is where the individual's WTA is bracketed between the highest dollar amount the respondent rejected in favor of the art print ($x) and the lowest amount they would accept ($z), where $x<$z; and in category (c), the individual prefers the art print to the highest dollar amount, which was $295 in this experiment. Assuming that the signed art print was not repulsive to the individual, response category (a) is bracketed from below by zero (i.e., if offered the print or zero dollars, they would take the print) and by $4. This bracketing along the real number line is illustrated below.



Response category (c), where the respondent states she would not accept the highest dollar amount offered over the good, does not provide an upper bound on the individual's WTA. However, we do know, with probability=1, that the respondent's WTA is larger than the upper dollar amount. Welsh and Bishop (339-40) use this observation to program the log likelihood function for this response category.

Using a multiple bounded approach to calculate a sample average WTA involves summing the estimated probability density function over the interval where the individual's response lies. The log likelihood function is:

$$[4.3] \quad \ln(Likelihood) = \sum_{1}^{n} \ln(Pi_{(\$x)} - Pi_{(\$z)}$$

where, $Pi_{(\$x)}$ and $Pi_{(\$z)}$ are the probabilities that respondent *i* would reject *$x* and accept *$z*, respectively, and *n* is the number of respondents.

For ease in computing the log likelihood function, the probability density function of WTA is often assumed to be logistically distributed. The log likelihood function is maximized with respect to the vector of coefficients *(B)* explaining the pattern of responses observed using a Gauss program developed by Welsh and Bishop. One of the variables must be the dollar amount the individual is asked to accept. Additional variables may include responses to attitude questions or the respondent's demographic characteristics, such as age and education. The vector of the first order conditions for maximum likelihood is shown in equation [4.4]:

$$[4.4] \quad \frac{\partial \ln(Likelihood)}{\partial B} = \sum_{i=1}^{n} \frac{1}{Pi_{(\$x)} - Pi_{(\$z)}} * \left[\frac{\partial Pi(\$x)}{\partial B} - \frac{\partial Pi(\$z)}{\partial B} \right] = 0$$

Using the coefficients estimated in equation [4.4], mean and median WTA can be calculated from equations [3.9] and [3.10], respectively.

The parametric approach is used in this paper for an important reason. Actual cash WTP is calculated from a binary choice logit model. Therefore, to provide comparability between distributional assumptions when comparing WTP derived using CVM and WTA derived using the method of PC, a logit based parametric approach is adopted for the PC analysis.

Testing for Differences Between WTP and WTA.

The equality of WTP and WTA is tested using three independent samples. To compare WTP^{DC} and WTA^{PC}, we estimate and compare confidence intervals based on the approach of Park et al. (1991). If the confidence intervals do not overlap, we conclude that WTA and WTP are statistically different.

ETIMATED EQUATIONS

Comparison of Demographics Across Sections.

Before testing for treatment effects, it is necessary to test whether the samples were significantly different or not in terms of standard demographics such as age, education, and income. To test for this across our three treatments, we performed one-way ANOVA's for education (F=0.74, p=0.48), age (F=2.89, p=0.06) and income (F=0.88, p=0.42). As indicated by the p values, the samples are not significantly different at the 0.1 level, although age would be significantly different at the 0.1 level.

Binary Logit Equations for CVM $WTP^{DC}(h:reminder)$ and $WTP^{DC}(a)$.

Equations [3.19] and [3.20] are replicated here as equations [4.5] and [4.6], representing the logit equations from the $WTP^{DC}(h:reminder)$ and $WTP^{DC}(a)$CVM treatments, respectively.

[4.5] $YPAY^{DC}(h)=-10.77-0.2578(\$BID)+1.96(MARKET)+7.84(LIKE)-0.537(AGE)+0.09(INC)$
(t statistics) (-1.75) (-2.38) (1.85) (2.27) (-2.12) (1.55)

[4.6] $YPAY^{DC}(a) =-7.92-0.1787(\$BID)+1.44(MARKET)+1.37(LIKE)-0.04(AGE)+0.05(INC)$
(t statistics) (2.05)(-2.56) (2.47) (1.88) (-0.88) (1.36)

Multiple Bounded Logit Equations for $WTA^{PC}(h)$.

The same basic variable specification was used to analyze the PC data using the multiple bounded logit model. The specification included the dollar amount they were asked

to accept *($BID)*, as well as income *(INC)*, *AGE* and *MARKET*. Due to the way the multiple bounded logit model is programmed, the dependent variable is coded as 0 if the respondent did not choose the print and 1 if the respondent chose to accept the print instead of the dollar amount. Thus, as the dollar amount offered rises, the odds of accepting the print goes down. Individuals whose responses to the series of *($BIDS)* contained circular triads (i.e., they agreed to accept a low amount of money instead of the art print and yet when offered a higher amount of money, chose the art print) were dropped from the analysis.

Loomis et al (1998:512) explain that "these circular triads may arise because the higher amount was offered first and then the lower amount, or simply because the point of indifference between money and the art print had been reached causing the individual to randomly switch choices". In any case, the multiple bounded likelihood function cannot handle such inconsistencies, as the individual appears simultaneously in several bid intervals, rather than just one. There were 24 individuals out of 103 responses, or about 23% of the participants, that had circular triads for the art print.

Although Loomis et al (1998:505) claimed that psychologists and probability theoreticians understood that this apparent transitivity may be random or systematic, implying that they could distinguish between circular triads that were truly intransitive and those which resulted from violation of the completeness axiom in the experimental design, there was no effort to make such a distinction regarding the circular triads. Rather, a full 23% of the sample were simply dropped from the analysis. Although psychologists may consider this more "efficient" than requiring larger sample sizes to begin with in order to accommodate the knowledge that many people will be indifferent between any two presented goods, the very definition of efficiency indicates that accuracy or competency is achieved

at the least cost. However, accuracy and competency were not achieved here, although they would not have been difficult to achieve. It is the absence of the fundamental economic axiom of completeness in this experimental design that is responsible for the circular triads and apparent intransitivity.

Equation [4.7] provides the coefficients and *t* statistics of the multiple bounded logit equation for the $WTA^{PC}(h)$.

[4.7] $ACCEPT\ PRINT=-2.47-0.0285(\$BID)+0.393(MARKET)+0.078(AGE)-0.015(INC)$
(*t* statistics) (-2.63)(-9.11) (3.84) (3.21) (-1.84)

The higher the dollar amount offered, the less likely the individual would accept the print instead of the money. Note that INC is also not a significant variable here. Although insignificant, the negative coefficient implies that the lower a respondent's income, the more likely he was to accept the print over the money.

RESULTS OF HYPOTHESES TESTS

H_0: $WTP^{DC}(a)=WTA^{PC}(h)$.

Table 4.1 presents the means, medians, and 95% confidence intervals for $WTP^{DC}(h:reminder)$, $WTP^{DC}(a)$, and $WTA^{PC}(h)$. In terms of our null hypothesis [4.1], mean $WTA^{PC}(h)$ exceeds $WTP^{DC}(a)$ by a factor of 5. The non-overlapping confidence intervals suggest these differences are significantly different at conventional levels suggesting rejection of null hypothesis [4.1]. Since this ratio encompasses both differences between hypothetical and real commitments as well as WTA vs. WTP, it is not surprising that it is large. Nonetheless, this ratio is smaller than many ratios of either hypothetical/actual or WTA/WTP found in the literature. In particular, the Welsh (1986:153) study is one of the few that elicits both hypothetical and actual WTA *and* WTP using the DC question format.

He found $WTA^{DC}(h)/WTP^{DC}(a)$ to be 14 in the 1984 Sandhill deer hunting permit DC experiments. However, it must be noted that deer hunting incorporates some public good characteristics (A hunter is permitted to privatize a public good.), as well as sport or recreation; whereas our good is strictly a private good, and income effects are expected to be greater for public goods. Our reduction in the ratio of WTA/WTP is similar to what Franciosi et al. found in their endowment experiments when individuals were endowed with the right to choose either a coffee mug or money.

Table 4.1.

Hypothetical and Actual WTP from CVM and Hypothetical WTA from PC

Treatment	Sample Size	Mean (Median) WTP or WTA by Treatment [95% CI of Mean]		
		WTP		WTA
		Dichotomous Choice 125		Paired Comparison
		Hypothetical Payment	Real Payment	Hypothetical MB Logit
No. 4	52	31(28) [20-37]		
No. 5	55		12 (9) [6-22]	
No. 6	79			59(52) [39-66]

H_0: WTP^{DC}(h:reminder)=WTA^{PC}(h).

In terms of our null hypothesis [4.2], mean $WTA^{PC}(h)$ is \$59 while mean $WTP^{DC}(h:reminder)$ is \$31. The non-overlapping 95% confidence intervals suggest we reject null hypothesis that the two are equal. The ratio of mean $WTA^{PC}(h)$ to mean $WTP^{DC}(h:reminder)$ is 1.9. This result is identical to what was obtained using a

nonparametric approach for $WTA^{PC}(h)$.[13] The difference in the median $WTA^{PC}(h)$ and median $WTP^{DC}(h{:}reminder)$ are smaller. Thus, the disparity is reduced when comparing the median.[14]

While the ratio of mean $WTA^{PC}(h)$ to mean $WTP^{DC}(h{:}reminder)$ is 1.9, this ratio is less than those reported in most other studies. As summarized by Kahnemann et al. (1990:1327), hypothetical mean WTA/hypothetical mean WTP is in the range of 2.6-16, averaging 5.38 across the seven studies cited. Our (median WTA/median WTP) is 1.85 as compared to 3.5 in the three studies cited by Kahnemann et al.

Our ratio of mean $WTA^{PC}(h)$ to mean $WTP^{DC}(h{:}reminder)$ is about the same value as the ratio other researchers find using deliverable goods and actual cash. In particular, Boyce et al. (1992) found the ratio of WTA(cash)/WTP(cash) to range from 1.66 to 2.36. Kahnemann et al.'s summary of four previous cash experiments had an average ratio of 4.5 for WTA(cash)/WTP(cash). Fisher et al. (1988), however, present a range of 1.6-16.6, including sixteen comparisons, although they do not specify whether these ratios involve means or medians, private goods or public goods, or hypothetical or cash scenarios.

DISCUSSION AND CONCLUSIONS

The problems presented by the failure to include indifference between two goods presented as an option make it very difficult to discuss the findings of this experiment in an economic context. Since WTA is a compensated demand function, and demand functions are

[13] Taken on face value, the number of rejections for each dollar amount describes the empirical (sample) WTA demand function for the art print. The sample mean of this empirical distribution is $59, identical to the logit estimate.

[14] Using the same empirical distribution calculated in footnote 6, the sample median is $39. The median is lower than the median estimated from the logit model because the empirical distributions are skewed.

derived from indifference curves, but the method of PC does permit an option of indifference, then it is clearly erroneous to claim that the measures derived herein are valid measures of WTA. The measures may have meaning within a psychological context, but they do not within an economic context. Therefore, it was inappropriate for Loomis et al. (1998:514) to claim that "[t]his study presents an alternative approach to estimate an individual's WTA." In any case, the study did not succeed in its objective. The chooser reference point did not avoid the loss aversion or endowment effects proposed by Kahnemann et al.

Although psychologists may find it interesting to force individuals to make a choice between two goods that they are indifferent between, microeconomic measures of WTA require the option of indifference. WTA is an economic measure, not a psychology measure. In economics, rational choice means that individuals have the option of indifference. Although psychologists may have a lot of insight helpful to the economics profession, it probably does not include simply disregarding all of the fundamental axioms of rational choice developed by the economics profession without first disproving the validity of those axioms; nor does it include the application of a methodology that has been developed in psychology to economics without ensuring consistency with economic theory, which consistency would not have been difficult to achieve. Thus, although there is no evidence that the circular triads in this experiment are "the product of inconsistent choice," for such a determination requires completeness, there is sufficient evidence that the method of PC itself is inconsistent with all the fundamental microeconomic axioms of rational choice and therefore microeconomic theory.

An alternative explanation for the disparities between WTA and WTP in private goods experiments is that the respondents are not likely to be in the market for the private goods used in these laboratory experiments. Microeconomic theory predicts the equality of WTA and WTP only at market clearing prices, but if respondents are not in the market for the private goods used in these laboratory experiments, then there is no reason to expect WTP to equal WTA. In Lucero (Working Paper), a reanalysis of the empirical evidence regarding WTP and WTA in private goods experiments indicates that all of these experiments were plagued by the fact that the respondents were not likely to be in the market for the private experimental goods and by the fact that the respondents in WTA treatments were consistently given a free good, which would only exacerbate the effects of respondents not being in the market to begin with. The reanalysis of these experiments indicated that both WTA and WTP were usually significantly below the actual market price of the good. An experiment designed to eliminate these two experimental design flaws is reported in Lucero, where it is found that when it can be ascertained that respondents are in the market for the experimental good and no one receives a free good, then WTA=WTP for relatively inexpensive private goods. The effect of respondents being in the market for art prints or not being in the market for art prints on hypothetical bias in our first experiment is presented in the next chapter.

CHAPTER FIVE

THE EFFECTS OF RESPONDENTS' BEING IN THE MARKET OR NOT BEING IN THE MARKET FOR PRIVATE EXPERIMENTAL GOODS ON HYPOTHETICAL BIAS

INTRODUCTION

It appears that there is more to the hypothetical bias than has been explained by the data analyses presented in this dissertation thus far. However, there is more information in the data that has not yet been analyzed that may help to explain the hypothetical bias reflected in the first experiment reported in this dissertation. We collected information on whether the respondents were in the *MARKET* for art prints on a five-point Likert scale, and *MARKET* was a significant determinant of WTP in all five treatments in the first experiment. Given that *MARKET* is a significant determinant of WTP, which is consistent with microeconomic predictions, then there should be a significant difference between the WTP given by those who reported that they were in the *MARKET* for art prints and those who reported that they were *NOT IN THE MARKET.*

In this chapter, an independent analysis of my proposal that WTP and hypothetical bias are strongly affected by the extent to which respondents are in the market or not in the market for private experimental goods is presented. This involves a two-part analysis. First, the three original independent samples from treatments No. 1 – 3 and presented in Chapter Two are divided according to those who either agreed or strongly agreed with the statement that they were in the market for art prints (*MARKET*) and those who neither agreed nor disagreed, disagreed, or strongly disagreed with the statement that they were in the market for art prints (*OTHERS*). These two sub-samples are then compared with the objective of determining whether there are significant differences in *WTP(h:reminder)*, *WTP(h:no reminder)*, and *WTP(a)* between them. These two sub-samples are then further compared in a re-testing of null hypotheses [2.1] – [2.7] presented in Chapter Two but with the original samples again divided between those

respondents in the *MARKET* for art prints and the *OTHERS* with the objective of testing for differences in hypothetical bias for these two groups and determining whether the reminder statement was effective in reducing or eliminating hypothetical bias. Lastly LAD regressions based on the paired samples are compared for those in the *MARKET* for art prints and the *OTHERS* with the objective of identifying significant differences in hypothetical and actual WTP and hypothetical bias for the two sub-samples.

Then those who neither agreed nor disagreed with the statement that they were in the market for art prints were excluded from both sub-samples, since they really do not belong in either, and since it is really a comparison of those who are strictly in the *MARKET* and those who are *"NOT IN THE MARKET"* for art prints that is relevant to market/non-market comparisons of WTP and hypothetical bias. Then all of the comparisons between those in the *MARKET* for art prints and the *OTHERS* were repeated for those in the *MARKET* and those *NOT IN THE MARKET* for art prints. Thus all of the null hypotheses presented in this chapter are tested twice – once when WTP behavior for those in the *MARKET* and the *OTHERS* is compared and twice when WTP behavior for those in the MARKET and those *NOT IN THE MARKET* is compared. Conducting both comparisons allows us to identify distinct patterns of WTP and hypothetical bias behavior for all three groups of respondents – those in the market for art prints, those not in the market for art prints, and those who neither agreed nor disagreed that they were in the market for art prints.

Lastly the dichotomous choice YES/NO contingency tables from treatments No. 4 and 5 (Table 1.1) and presented in Chapter Three are also divided in two according to those who were in the *MARKET* for art prints and those who were *NOT IN THE*

MARKET for art prints. Although tests for differences in the contingency tables are not run, a visual inspection clearly indicates that the WTP behavior in the hypothetical and actual cash dichotomous choice treatments confirm the findings with respect to the open-ended treatments.

RESEARCH DESIGN

Null hypotheses [5.1] – [5.3] simply test whether there is a significant difference in the WTP measures elicited in treatments No. 1, 2, and 3, respectively, for those respondents who indicated that they were in the *MARKET* for art prints (those who either agreed or strongly agreed with the statement that they were in the *MARKET* for art prints) and all of the *OTHER* respondents in the independent sample (including those who neither agreed nor disagreed, disagreed, or strongly disagreed with the statement that they were in the *MARKET* for art prints). The remaining null hypotheses simply reproduce null hypotheses [2.1] – [2.7] with the original sample size divided between those who indicated that they were in the *MARKET* for art prints and all the *OTHERS* remaining in the original sample.

Subsequent to this initial analysis, those who neither agreed nor disagreed with the statement that they were in the *MARKET* for art prints are excluded from both samples. They really do not belong with either the group who is in the *MARKET* or the group who is *NOT IN THE MARKET.* Thus all of the null hypotheses presented in this chapter are tested twice. The first group of tests compare those in the *MARKET* with the *OTHERS,* and the second group of tests compare those in the *MARKET* with those *NOT IN THE MARKET.* Although it is this latter comparison that is truly relevant to market/non-market comparisons, the first analysis is also presented because the

difference between the two comparisons produces some very definitive and significant results pertaining to the behavior of all three groups of individuals - those who were in the *MARKET* for art prints, those who were *NOT IN THE MARKET* for art prints, and those who were neither in the *MARKET* or *NOT IN THE MARKET* for art prints.

Treatments.

The treatments compared and discussed in this chapter are as follows:

No. 1: *WTP(h:no reminder)*: hypothetical WTP asked as in a standard CVM survey, using wording similar to Neill et al.

No. 2: *WTP(h:reminder)*: hypothetical WTP asked after subjects are reminded not to give what they think a fair price is, or what it sells for, and to act as if they were in a real market with their real budget.

No. 3: *WTP(a)*: actual WTP in the form of cash, check, or promissory note.

No. 4: $WTP^{DC}(h{:}reminder)$: hypothetical DC WTP with "reminder;" and

No. 5: $WTP^{DC}(a)$: actual cash DC WTP.

Null Hypotheses.

Independent Samples.

[5.1] $H_0{:}WTP^{OE}(h{:}no\ reminder)_M = WTP^{OE}(h{:}no\ reminder)_O$;

[5.2] $H_0{:}WTP^{OE}(h{:}reminder)_M = WTP^{OE}(h{:}reminder)_O$;

[5.3] $H_0{:}WTP^{OE}(a)_M = WTP^{OE}(a)_O$;

[5.4] $H_0{:}WTP^{OE}(a)_M = WTP^{OE}(h{:}no\ reminder)_M$;

[5.5] $H_0{:}WTP^{OE}(a)_O = WTP^{OE}(h{:}no\ reminder)_O$;

[5.6] $H_0{:}WTP^{OE}(a)_M = WTP^{OE}(h{:}reminder)_M$;

[5.7] $H_0{:}WTP^{OE}(a)_O = WTP^{OE}(h{:}reminder)_O$;

[5.8] $H_0{:}WTP^{OE}(h{:}no\ reminder)_M = WTP^{OE}(h{:}reminder)_M$; and

[5.9] $H_0{:}WTP^{OE}(h{:}no\ reminder)_O = WTP^{OE}(h{:}reminder)_O$

where the subscript *M* indicates that the sample includes only those respondents who indicated that they were in the *MARKET* for art prints; and where the subscript *O* indicates that the sample includes all of the *OTHER* respondents in the independent sample.

Paired Samples.

[5.10] $H_0: WTPi^{OE}(a)_M = WTPi^{OE}(h:\text{no reminder})_M$;

[5.11] $H_0: WTPi^{OE}(a)_O = WTPi^{OE}(h:\text{no reminder})_O$;

[5.12] $H_0: WTP_j^{OE}(a)_M = WTP_j^{OE}(h:\text{reminder})_M$;

[5.13] $H_0: WTP_j^{OE}(a)_O = WTP_j^{OE}(h:\text{reminder})_O$;

[5.14] $H_0: WTPi^{OE}(a)_M = B_0 + B_1[WTPi^{OE}(h:\text{no reminder})_M]$;

[5.15] $H_0: WTPi^{OE}(a)_O = B_0 + B_1[WTPi^{OE}(h:\text{no reminder})_O]$;

[5.16] $H_0: WTP_j^{OE}(a)_M = B_0 + B_1[WTP_j^{OE}(h:\text{reminder})_M]$; and

[5.17] $H_0: WTP_j^{OE}(a)_O = B_0 + B_1[WTP_j^{OE}(h:\text{reminder})_O]$

where the subscript *i* indicates paired samples from treatment No. 1 and the subscript *j* indicates paired samples from treatment No. 2.

DISCUSSION OF STATISTICAL TESTS AND DESCRIPTIVE STATISTICS

The Multi-Response Permutation Procedure (MRPP) and Permutation Tests for Matched Pairs (PTMP), both discussed in detail in Chapter Two, are the only tests of equality reported in this chapter, for they are the most appropriate tests for our data since they are the only truly non-parametric tests encountered thus far. Given that Mielke (Slauson 1991:3) and Hanemann (1984:339) advocate tests for the equality of medians over tests for the equality of means, only median tests are reported here (V=1). Further, given that degrees of freedom are irrelevant to permutations, the intragroup average

distances are weighted by relative sample size (C=1). Thus only the V=1 and C=1 options are reported for both MRPP and PTMP. Least absolute deviation (discussed in the DISCUSSION AND CONCLUSIONS section of Chapter Three) is also used instead of ordinary least squares regression to test null hypotheses [5.14] – [5.17].

The descriptive statistics for the six samples that result from dividing the original three independent treatments No. 1, 2, and 3 in two, representing those who reported that they were in the *MARKET* for art prints and all *OTHERS*, are reported in Table 5.1.

Table 5.1.

Descriptive Statistics:
MARKET/OTHER Comparisons of Independent Samples

Sample	Count	Mean	Median	Mode	Range
$WTP^{OE}(h{:}no\ reminder)_M$	9	95.00	50.00	45	15-400
$WTP^{OE}(h{:}no\ reminder)_O$	26	24.11	20.00	5	0-75
$WTP^{OE}(h{:}reminder)_M$	9	42.22	35.00	20	15-100
$WTP^{OE}(h{:}reminder)_O$	23	19.46	10.00	0	0-75
$WTP^{OE}(a)_M$	10	20.48	19.90	N/A	0-47
$WTP^{OE}(a)_O$	22	11.75	10.00	5	0-50

The descriptive statistics for the paired samples are presented in Table 5.2. Both Tables 5.1 and 5.2 reflect that the number of respondents who reported that they were in the *MARKET* for art prints are consistently fewer than the number of *OTHERS*. They also reflect that the median, mode, and range are consistently greater for those respondents who were in the *MARKET*. The sub-samples for the *OTHERS* contained a lot of zeros and low values. With the *reminder*, more than half of the *OTHERS* bid $10 or less. There were no zero bids for those in the *MARKET* either with or with *no reminder*. For those in the *MARKET*, there was only one zero bid in the actual cash treatment in the independent samples. That person disagreed that he or she would *BUY* the art and neither agreed nor disagreed that he *LIKED The Shadow Hunter* or that he or she would buy it as a *GIFT*.

Table 5.2.

Descriptive Statistics:
MARKET/OTHER Comparisons of Paired Samples

Sample	Mean	Median	Mode	Range
$WTP_i^{OE}(h{:}no\ reminder)_M$	95.00	50	45	15-400
$WTP_i^{OE}(a)_M$	23.67	25	25	8-40
$WTP_i^{OE}(h{:}no\ reminder)_O$	24.11	20	5	0-75
$WTP_i^{OE}(a)_O$	7.46	5	0	0-30
$WTP_i^{OE}(h{:}reminder)_M$	42.22	35	20	15-100
$WTP_i^{OE}(a)_M$	27.50	15	10	5-100
$WTP_i^{OE}(h{:}reminder)_O$	19.46	10	0	0-75
$WTP_i^{OE}(a)_O$	8.27	5	0	0-50

Table 5.3 reports both mean and median hypothetical/actual cash WTP ratios for both those in the *MARKET* and the *OTHERS* for the independent samples. It indicates that with independent samples, the hypothetical/actual cash WTP ratio is consistently greater for those in the *MARKET*. The median ratio is notably 1.00 with the *reminder* for the *OTHERS*.

Table 5.3.

Measures of Hypothetical Bias in Independent Samples
MARKET/OTHER Comparisons

Ratio	Means	Medians
$[WTP^{OE}(h{:}no\ reminder)_M]/[WTP^{OE}(a)_M]$	4.64	2.51
$[WTP^{OE}(h{:}no\ reminder)_O]/[WTP^{OE}(a)_O]$	2.05	2.00
$[WTP^{OE}(h{:}reminder)_M]/[WTP^{OE}(a)_M]$	2.06	1.76
$[WTP^{OE}(h{:}reminder)_O]/[WTP^{OE}(a)_O]$	1.66	1.00

Table 5.4 reports both mean and median hypothetical/actual cash WTP ratios for both those in the *MARKET* and the *OTHERS* for the paired samples. There is no clear pattern in a comparison of the ratios for those in the *MARKET* and the *OTHERS*.

Table 5.4.

Measures of Hypothetical Bias in Paired Samples
MARKET/OTHER Comparisons

Ratio	Means	Medians
$[WTP_i^{OE}(h{:}no\ reminder)_M]/[WTP_i^{OE}(a)_M]$	4.01	2.00
$[WTP_i^{OE}(h{:}no\ reminder)_O]/[WTP_i^{OE}(a)_O]$	3.23	4.00
$[WTP_i^{OE}(h{:}reminder)_M]/[WTP_i^{OE}(a)_M]$	1.54	2.33
$[WTP_i^{OE}(h{:}reminder)_O]/[WTP_i^{OE}(a)_O]$	2.35	2.00

RESULTS OF FORMAL HYPOTHESES TESTS
MARKET/OTHER COMPARISONS

Independent Samples.

The MRPP test results of null hypotheses [5.1] – [5.9] are reported in Table 5.5.

Table 5.5.

MRPP Tests for the Equality of MARKET and OTHER Medians

Null Hypotheses	*p* values
$H_0{:}WTP^{OE}(h{:}no\ reminder)_M=WTP^{OE}(h{:}no\ reminder)_O$	0.00075
$H_0{:}WTP^{OE}(h{:}reminder)_M=WTP^{OE}(h{:}reminder)_O$	0.025
$H_0{:}WTP^{OE}(a)_M=WTP^{OE}(a)_O$	0.08722
$H_0{:}WTP^{OE}(a)_M=WTP^{OE}(h{:}no\ reminder)_M$	0.00206
$H_0{:}WTP^{OE}(a)_O=WTP^{OE}(h{:}no\ reminder)_O$	0.00635
$H_0{:}WTP^{OE}(a)_M=WTP^{OE}(h{:}reminder)_M$	0.08337
$H_0{:}WTP^{OE}(a)_O=WTP^{OE}(h{:}reminder)_O$	0.20236
$H_0{:}WTP^{OE}(h{:}no\ reminder)_M=WTP^{OE}(h{:}reminder)_M$	0.28117
$H_0{:}WTP^{OE}(h{:}no\ reminder)_O=WTP^{OE}(h{:}reminder)_O$	0.20845

The test result for the first null hypothesis indicates that with *no reminder*, there is a significant difference in hypothetical WTP between those who reported being in the *MARKET* and the *OTHERS* at the 0.01 level of significance. The *p* value for the second null hypothesis indicates that with the *reminder*, we can reject equality of hypothetical WTP between those who reported being in the *MARKET* and the *OTHERS* at the 0.025 level of significance. The *p* value for the third null hypothesis indicates that we cannot reject equality of actual cash WTP between those who reported being in the *MARKET*

and the *OTHERS* at the 0.05 level of significance. This is consistent with the findings in Chapter Two – thus far, we cannot reject equality of any measure of actual cash WTP with any other measure of actual cash WTP at the 0.05 level of significance or better.

The *p* values for the fourth and fifth null hypotheses in Table 5.5 indicate that with *no reminder*, we reject equality of actual and hypothetical WTP for those who were in the *MARKET* and the *OTHERS*. The *p* values for the eighth and ninth null hypotheses in Table 5.5 indicate that we cannot reject equality of hypothetical WTP with and with *no reminder* for either those who were in the *MARKET* or the *OTHERS*.

The test results for the sixth and seventh null hypotheses indicate that with the *reminder*, we cannot reject equality of hypothetical and actual WTP for either those in the *MARKET* or the *OTHERS* at the 0.05 level of significance. This is inconsistent with the findings for the undivided comparison. The *p* value yielded by MRPP when V=1 and C=1 for the equality of the medians of the undivided independent samples from treatments No. 2 and 3 is 0.04523, causing us to reject equality at the .05 level of significance. The reason for these seemingly inconsistent results is probably due to the fact that the divided samples are so small. As with most statistical tests, permutation tests gain power with increasing sample size. Exact MRPP tests were not run for either the divided or undivided samples because this requires computer capabilities not presently available to me. Perhaps with larger sample sizes equality would also be rejected for the divided samples. Exact *p* values might also change any or all of the findings presented in this chapter, as well as those related to MRPP tests presented in Chapters Two and Three.

Paired Samples.

The PTMP test results of null hypotheses [5.10] – [5.13] are reported in Table 5.6.

Table 5.6.

PTMP Tests for the Equality of MARKET and OTHER Medians

Null Hypotheses	*p* values
H_0: $WTP_i^{OE}(a)_M = WTP_i^{OE}(h\text{:no reminder})_M$	0.00390
H_0: $WTP_i^{OE}(a)_O = WTP_i^{OE}(h\text{:no reminder})_O$	0.00000
H_0: $WTP_i^{OE}(a)_M = WTP_i^{OE}(h\text{:reminder})_M$	0.06250
H_0: $WTP_i^{OE}(a)_O = WTP_i^{OE}(h\text{:reminder})_O$	0.00097

These results again reflect that with *no reminder*, we reject equality of hypothetical and actual WTP for those in the *MARKET* and the *OTHERS*. This is the only case in Tables 5.3 and 5.4 where the median hypothetical/actual cash WTP ratio is smaller for those in the *MARKET* than for the *OTHERS*. With no *reminder*, it is 2:1 for those in the *MARKET* and 4:1 for the *OTHERS*.

The *p* values for the twelfth and thirteenth null hypotheses indicate that with the *reminder*, we cannot reject equality of hypothetical and actual cash WTP for those in the *MARKET*, but we do reject equality of hypothetical and actual cash WTP for the *OTHERS* at the 0.01 level of significance. This seems odd, since the median hypothetical/actual cash WTP ratio is smaller for the *OTHERS* than for those in the *MARKET* (although the mean ratio is smaller for those in the *MARKET* than for the *OTHERS*). However, a review of the raw data indicates that the largest hypothetical/actual cash WTP was 7.5:1 for those in the *MARKET* and 30:1 for the *OTHERS*.

$H_0: WTP_i^{OE}(a)_M = B_0 + B_1[WTP_i^{OE}(h:no\ reminder)_M]$

An LAD regression of this null hypothesis generates B_0=25 and B_1=0. This indicates that with *no reminder*, the respondents in the *MARKET* were most likely to bid $25, but beyond that there was no significant relationship between their hypothetical and actual cash WTP. The median and mode for this paired sample measure of actual WTP is $25, and the mean is $23.67. The *p* value of the full model is 0.85320.

$H_0: WTP_i^{OE}(a)_O = B_0 + B_1[WTP_i^{OE}(h:no\ reminder)_O]$

An LAD regression of this null hypothesis generates B_0=0 and B_1=.20. This indicates that with *no reminder*, the *OTHERS* were most likely to bid zero, and beyond that each dollar of their hypothetical WTP translated into only $.20 of actual cash WTP.

$H_0: WTP_i^{OE}(a)_M = B_0 + B_1[WTP_i^{OE}(h:reminder)_M]$

An LAD regression of this null hypothesis generates B_0=-10 and B_1=1. This indicates that with the *reminder*, there was a one to one correspondence between hypothetical and actual cash WTP minus $10 for those respondents in the *MARKET.* Three of the nine respondents bid the same amount in the hypothetical and actual cash treatments, and one respondent even strategically bid slightly more in the actual cash treatment. The largest hypothetical/actual cash WTP ratio for any individual was 7.5:1. The *p* value of the full model is 0.06759.

$H_0: WTP_i^{OE}(a)_O = B_0 + B_1[WTP_i^{OE}(h:reminder)_O]$

An LAD regression of this null hypothesis generates B_0=0 and B_1=.4228. This indicates that with *no* reminder, the *OTHERS* were most likely to bid zero, and after that their hypothetical WTP translated into $.42 of actual cash WTP. Of the 23 respondents in this sub-sample, 11 bid the same in both treatments. All but one of these 11 respondents

bid $10 or less. The other of these eleven bid $25 in both the hypothetical and actual cash WTP treatments. This person neither agreed nor disagreed that he or she was in the *MARKET* for art prints. No one bid more in the actual cash treatment. The largest hypothetical/actual cash WTP ratio for any individual was 30:1.

FINDINGS AND CONCLUSIONS PERTAINING TO MARKET/OTHER COMPARISONS

The independent sample analyses indicate that hypothetical WTP is significantly different for those in the *MARKET* and the *OTHERS* both with and with *no reminder* at the 0.05 level of significance. However, actual cash WTP is not significantly different at the 0.05 level of significance. The independent sample analyses further indicate that with *no reminder*, equality of hypothetical and actual cash WTP is rejected for both those in the *MARKET* and the *OTHERS*. This is consistent with the paired sample tests for equality of the medians and LAD regression with *no reminder*.

The independent and paired sample analyses with the *reminder*, however, are not entirely consistent. The independent sample analyses indicate that we cannot reject median equality of hypothetical and actual cash WTP for both those in the *MARKET* and the *OTHERS* at the 0.05 level of significance. The median hypothetical/actual cash WTP ratio of 1:1 for the *OTHERS* indicates that the *reminder* was extremely effective in eliminating the hypothetical bias for this group of people. The fact that we also cannot reject equality of median hypothetical and actual cash WTP with independent samples for those in the *MARKET*, even though their median hypothetical/actual cash WTP ratio was 1.76:1 indicates that this measure of hypothetical bias does not necessarily need to be 1:1 or very close for a finding of no significant difference. The finding of no significant difference for those in the *MARKET* indicates that the *reminder* was also very effective in

eliminating hypothetical bias for this group of people. Division of the original independent samples into those who were in the *MARKET* and the *OTHERS* produces much stronger evidence that the *reminder* accomplished its objective of eliminating any significant hypothetical bias than did analysis of the undivided samples in Chapter Two, for MRPP does not reject equality of these divided samples as it rejects equality of the undivided samples. This finding of no significant difference is supported by the Kolmogorov-Smirnov test of the distributions of the undivided samples and the Mann-Whitney test of medians of the undivided samples.

The paired sample analyses indicate that with the *reminder*, we cannot reject equality of hypothetical and actual cash WTP for those in the *MARKET* at the 0.05 level of significance, but we do reject equality for the *OTHERS* at the 0.01 level of significance. The LAD regressions with the *reminder* indicate that B_1=1 for those in the *MARKET* and B_1=0.42 for the *OTHERS*. The largest hypothetical/actual cash WTP ratio for the individuals in the *MARKET* was 7.5:1, whereas it was 30:1 for the *OTHERS*. The fact that the median hypothetical/actual cash WTP ratio was 2.33 for those in the *MARKET* and 2.00 for the *OTHERS* seems to indicate that there is more to measuring equality than is reflected in this ratio, and that perhaps this ratio is not a very reliable measure of hypothetical bias. The mean ratio was 1.54 for those in the *MARKET* and 2.35 for the *OTHERS*. Thus for those in the *MARKET* we cannot reject equality of hypothetical and actual cash WTP in either the independent or paired sample analyses when the *reminder* is included. This is in spite of the fact that the paired sample included a $100 bid in both the hypothetical and actual cash treatments, but the actual cash independent WTP treatment did not.

The inconsistencies between the findings of the independent and paired samples just presented raise the issue of whether independent or paired samples are more appropriate in testing for the presence hypothetical bias. Is it the equality of one person's hypothetical and actual cash WTP or the equality of one person's hypothetical WTP with another person's actual cash WTP that is relevant in testing for hypothetical bias?

In this next section, those who neither agreed nor disagreed that they were in the market for art prints were completely excluded from both sub-samples. Thus the comparisons will be of those strictly in the *MARKET* and those strictly *NOT IN THE MARKET*. There were 7, 9, and 6 respondents who neither agreed nor disagreed that they were in the *MARKET* for art prints in treatments No. 1, 2, and 3, respectively, who were excluded from *OTHERS* to arrive at those *NOT IN THE MARKET*.

DESCRIPTIVE STATISTICS – MARKET/NON-MARKET COMPARISONS

The descriptive statistics for the six samples representing those who were in the *MARKET* for art prints and those who were *NOT IN THE MARKET* for art prints in treatments No. 1, 2, and 3 are reported in Table 5.7. The subscript *N* reflects that the sample represents those *NOT IN THE MARKET* for art prints.

Table 5.7.

Descriptive Statistics:
Market/Non-Market Comparisons of Independent Samples

Sample	Count	Mean	Median	Mode	Range
$WTP^{OE}(h{:}no\ reminder)_M$	9	95.00	50.00	45	15-400
$WTP^{OE}(h{:}no\ reminder)_N$	19	24.05	20.00	20	0-75
$WTP^{OE}(h{:}reminder)_M$	9	42.22	35.00	20	15-100
$WTP^{OE}(h{:}reminder)_N$	14	17.79	7.50	0	0-75
$WTP^{OE}(a)_M$	10	20.48	19.90	N/A	0-47
$WTP^{OE}(a)_N$	16	8.97	8.75	5	0-30

A comparison of Tables 5.1 and 5.7 indicates that with *no reminder*, the mode is 20 for those *NOT IN THE MARKET* and 5 for the *OTHERS*. It further indicates that the means and medians for both hypothetical and actual cash WTP in treatments No. 2 and 3, respectively, were lower for those *NOT IN THE MARKET* than for the *OTHERS*.

The descriptive statistics for the paired samples are reported in Table 5.8. A comparison of Tables 5.2 and 5.8 indicates that the means for hypothetical and actual cash WTP were lower in treatment No. 1 for those *NOT IN THE MARKET* than for the *OTHERS*, but again the mode in the hypothetical treatment went from 5 to 20. The means and medians for both hypothetical and actual cash WTP are lower for those *NOT IN THE MARKET* than for the *OTHERS*, and the median notably went from 5 to 1.

Table 5.8.

Descriptive Statistics:
Market/Non-Market Comparisons of Paired Samples

Sample	Mean	Median	Mode	Range
$WTP_i^{OE}(h{:}no\ reminder)_M$	95.00	50.00	45	15-400
$WTP_i^{OE}(a)_M$	23.67	25.00	25	8-40
$WTP_i^{OE}(h{:}no\ reminder)_N$	24.05	20.00	20	0-75
$WTP_i^{OE}(a)_N$	6.26	5.00	0	0-30
$WTP_i^{OE}(h{:}reminder)_M$	42.22	35.00	20	15-100
$WTP_i^{OE}(a)_M$	27.50	15.00	10	5-100
$WTP_i^{OE}(h{:}reminder)_N$	17.79	7.50	0	0-75
$WTP_i^{OE}(a)_N$	6.71	1.00	0	0-50

Table 5.9 reports mean and median hypothetical/actual cash WTP ratios for those in the *MARKET* and those *NOT IN THE MARKET* for the independent samples. A comparison of Tables 5.3 and 5.9 indicates that with *no reminder*, the mean and median ratio measures of hypothetical bias are both greater for those *NOT IN THE MARKET* than for the *OTHERS*. The mean ratio is also greater with the *reminder*; however, the median ratio is notably smaller – it fell from 1.00 to .857. As reflected in Table 5.7, for those

NOT IN THE MARKET median actual cash WTP was greater than median hypothetical WTP with the *reminder*.

Table 5.9.

Measures of Hypothetical Bias in Independent Samples
Market/Non-Market Comparisons

Ratio	Means	Medians
$[WTP^{OE}(h\text{:no reminder})_M]/[WTP^{OE}(a)_M]$	4.64	2.51
$[WTP^{OE}(h\text{:no reminder})_N]/[WTP^{OE}(a)_N]$	2.68	2.29
$[WTP^{OE}(h\text{:reminder})_M]/[WTP^{OE}(a)_M]$	2.06	1.76
$[WTP^{OE}(h\text{:reminder})_N]/[WTP^{OE}(a)_N]$	1.98	.857

Table 5.10 reports mean and median hypothetical/actual cash WTP ratios for those in the *MARKET* for art prints and those *NOT IN THE MARKET*. A comparison of Tables 5.4 and 5.10 indicates that both with and with *no reminder* the mean ratio is slightly greater for those *NOT IN THE MARKET* than for the *OTHERS*; however, with the *reminder* the median ratio is 7.5 for those *NOT IN THE MARKET* compared with 2.00 for the *OTHERS*. This dramatic increase in the median ratio reflects the fact that even though hypothetical WTP is only 7.5 (well below the actual purchase price to us of $35 for the art print), actual cash WTP is only 1 for those *NOT IN THE MARKET* as compared to 5 for the *OTHERS* with a mode of 0 constant for both groups.

Table 5.10.

Measures of Hypothetical Bias in Paired Samples
Market/Non-Market Comparisons

Ratio	Means	Medians
$[WTP_i^{OE}(h\text{:no reminder})_M]/[WTP_i^{OE}(a)_M]$	4.01	2.00
$[WTP_i^{OE}(h\text{:no reminder})_N]/[WTP_i^{OE}(a)_N]$	3.84	4.00
$[WTP_i^{OE}(h\text{:reminder})_M]/[WTP_i^{OE}(a)_M]$	1.54	2.33
$[WTP_i^{OE}(h\text{:reminder})_N]/[WTP_i^{OE}(a)_N]$	2.65	7.5

RESULTS OF FORMAL HYPOTHESES TESTS MARKET/NON-MARKET COMPARISONS

Independent Samples.

Table 5.11 reports the MRPP test results of null hypotheses [5.1] – [5.9] where those *NOT IN THE MARKET* (denoted by the subscript *N*) have replaced the *OTHERS* in the null hypotheses.

A comparison of Tables 5.5 and 5.11 indicates that for the first time in this dissertation the equality of one measure of actual cash WTP with that of another can be rejected. The *p* value of 0.01929 for the third null hypothesis indicates that actual cash WTP for those in the *MARKET* is significantly different from actual cash WTP for those *NOT IN THE MARKET*. Table 5.7 reflects that the *MARKET/NON-MARKET* mean WTP ratio is 3.95, 2.37, and 2.28 for treatments No. 1, 2, and 3, respectively; and the *MARKET/NON-MARKET* median WTP ratio is 2.5, 4.67, and 2.27 for treatments No. 1, 2, and 3, respectively. Thus all three independent treatments confirm that those in the *MARKET* for art prints and those *NOT IN THE MARKET* come from two separate and distinct populations. Therefore, they should not be combined in one sample.

Table 5.11.

MRPP Tests for the Equality of Market and Non-Market Medians

Null Hypotheses	*p* values
H_0: $WTP^{OE}(h{:}no\ reminder)_M = WTP^{OE}(h{:}no\ reminder)_N$	0.00201
H_0: $WTP^{OE}(h{:}reminder)_M = WTP^{OE}(h{:}reminder)_N$	0.02667
H_0: $WTP^{OE}(a)_M = WTP^{OE}(a)_N$	0.01929
H_0: $WTP^{OE}(a)_M = WTP^{OE}(h{:}no\ reminder)_M$	0.00206
H_0: $WTP^{OE}(a)_N = WTP^{OE}(h{:}no\ reminder)_N$	0.00345
H_0: $WTP^{OE}(a)_M = WTP^{OE}(h{:}reminder)_M$	0.08337
H_0: $WTP^{OE}(a)_N = WTP^{OE}(h{:}reminder)_N$	0.14622
H_0: $WTP^{OE}(h{:}no\ reminder)_M = WTP^{OE}(h{:}reminder)_M$	0.28117
H_0: $WTP^{OE}(h{:}no\ reminder)_N = WTP^{OE}(h{:}reminder)_N$	0.18723

Paired Samples.

Table 5.12 reports the PTMP test results of null hypotheses [5.10] – [5.13] where those *NOT IN THE MARKET* (denoted by the subscript *N*) have replaced the *OTHERS* in the null hypotheses.

Table 5.12.

PTMP Tests for the Equality of Market/Non-Market Medians

Null Hypotheses	*p* values
$H_0: WTP_i^{OE}(a)_M = WTP_i^{OE}(h\text{:no reminder})_M$	0.00390
$H_0: WTP_i^{OE}(a)_N = WTP_i^{OE}(h\text{:no reminder})_N$	0.00024
$H_0: WTP_i^{OE}(a)_M = WTP_i^{OE}(h\text{:reminder})_M$	0.06250
$H_0: WTP_i^{OE}(a)_N = WTP_i^{OE}(h\text{:reminder})_N$	0.06250

A comparison of Tables 5.6 and 5.12 indicates that when those who neither agreed not disagreed that they were in the *MARKET* for art prints are excluded, then we cannot reject equality of hypothetical and actual cash WTP for those in the *MARKET* nor for those *NOT IN THE MARKET* with the *reminder*. For those *NOT IN THE MARKET*, even though the median ratio of hypothetical/actual cash WTP was 7.5:1, out of the 14 observations there were nine ties all for $10 or less, and one person even bid $2 more in the actual treatment than in the hypothetical treatment. This again indicates that perhaps hypothetical/actual cash WTP ratios may not be very good measures of hypothetical bias.

$$H_0: WTP_i^{OE}(a)_N = B_0 + B_1[WTP_i^{OE}(h\text{:no reminder})_N]$$

An LAD regression of this null hypothesis generates B_0=0 and B_1=.09. A comparison of this equation with the parallel equation for the *OTHERS*, which yielded B_0=0 and B_1=.20, indicates that with *no reminder* those *NOT IN THE MARKET* were also most likely to bid zero, but every dollar of their hypothetical WTP translates into only $.09. The *p* value for the full model is 0.62340.

$H_0: WTP_i^{OE}(a)_N = B_0 + B_1[WTP_i^{OE}(h{:}reminder)_N]$

An LAD regression of this null hypothesis generates B_0=0 and B_1=0.17241. A comparison of this equation with the parallel equation for the *OTHERS*, which yielded B_0=0 and B_1=0.4228, indicates that with the *reminder* those *NOT IN THE MARKET* were also most likely to bid zero, but every dollar of their hypothetical WTP translates into only $.17.The *p* value for the full model is 0.00680.

MARKET/NON-MARKET COMPARISONS OF DICHOTOMOUS CHOICE CONTINGENCY TABLES

A separation of the dichotomous choice contingency tables of NO responses presented in Chapter Three into those in the *MARKET* for art prints and those *NOT IN THE MARKET* for art prints are presented in Tables 5.13 and 5.14, respectively. An N/A in the tables indicates that *$BID* amount was not included in any the questionnaires received by the particular sample under comparison.

TABLE 5.13

CONTINGENCY TABLE OF NO RESPONSES FOR THOSE IN THE MARKET

Bid	$2	$4	$8	$12	$18	$24	$36	$48	$72	$120
NO in Hypothetical	N/A	0	0	0	0	N/A	2	0	2	1
NO in Actual	0	0	N/A	0	1	3	2	1	2	3

TABLE 5.14

CONTINGENCY TABLE OF NO RESPONSES FOR THOSE NOT IN THE MARKET

Bid	$2	$4	$8	$12	$18	$24	$36	$48	$72	$120
NO in Hypothetical	0	2	1	1	3	2	N/A	3	2	2
NO in Actual	1	1	3	4	1	1	3	3	1	N/A

A separation of the dichotomous choice contingency tables of YES responses presented in Chapter Three into those in the *MARKET* for art prints and those *NOT IN THE MARKET* for art prints are presented in Tables 5.15 and 5.16, respectively.

TABLE 5.15

CONTINGENCY TABLE OF YES RESPONSES FOR THOSE IN THE MARKET

Bid	$2	$4	$8	$12	$18	$24	$36	$48	$72	$120
YES in Hypothetical	N/A	1	1	1	1	N/A	2	1	0	0
YES in Actual	4	1	N/A	1	1	0	0	0	0	0

TABLE 5.16

CONTINGENCY TABLE OF YES RESPONSES FOR THOSE NOT IN THE MARKET

Bid	$2	$4	$8	$12	$18	$24	$36	$48	$72	$120
YES in Hypothetical	2	0	0	0	0	0	N/A	0	0	0
YES in Actual	1	0	0	1	0	0	0	0	0	N/A

Tests for the equality of the percentage of NO and YES responses in these contingency tables were not run, but they are not necessary to see that there is a significant difference in the percentage of NO (YES) responses received by those in the *MARKET* and those *NOT IN THE MARKET* for art prints. There was clearly a strong tendency for those in the *MARKET* to respond YES to lower *$BID* amounts and NO to higher *$BID* amounts; and for those *NOT IN THE MARKET* to respond NO to all *$BID* amounts. In the hypothetical treatment, there were only two YES responses – both at the $2 bid amount. In the actual cash treatment, there were also only two YES responses - one at the $2 bid amount and one at the $12 bid amount.

DISCUSSION AND CONCLUSIONS

The *MARKET/NON-MARKET* analyses of the independent samples presented herein indicate that *WTP(h:reminder)*, *WTP(h:no reminder)*, and *WTP(a)* are all

significantly greater for those in the *MARKET* for art prints than for those *NOT IN THE MARKET* at the 0.05 level of significance. The significance of this finding is reflected in the fact that this is the only comparison in which we can reject equality of one measure of actual cash WTP with another measure of actual cash WTP. Thus those in the *MARKET* for art prints and those *NOT IN THE MARKET* come from two separate and distinct populations with separate and distinct WTP behaviors. Therefore, it would be statistically inappropriate to combine them in one sample.

Combining those in the *MARKET* and those *NOT IN THE MARKET* leads to the skewness reflected in our samples. It is the WTP for those *NOT IN THE MARKET* that is characterized by many zero and low bids. Further, independent comparisons can be very misleading if the samples are not characterized by the same percentage of respondents in the *MARKET* or *NOT IN THE MARKET* for any private good used in any laboratory experiment. For example, if the hypothetical WTP treatment consisted mostly of respondents who were in the *MARKET* and the actual WTP treatment consisted mostly of respondents *NOT IN THE MARKET*, then any actual hypothetical bias present would be exaggerated. On the contrary, if the situation were reversed, then any hypothetical bias present would be severely underestimated, and we could end up with results where actual cash WTP is actually less than hypothetical WTP, as suggested by the .857 measure of hypothetical bias reflected in Table 5.9. Although tests for differences in the percentage of NO (YES) responses in the hypothetical and actual cash dichotomous choice treatments between those in the *MARKET* and those *NOT IN THE MARKET* for art prints were not run, a visual inspection of contingency Tables 5.13 – 5.16 clearly indicates that the dichotomous choice data confirms the findings from comparisons of the open-ended

data that WTP behavior is significantly different for those in the *MARKET* and those *NOT IN THE MARKET.*

The *MARKET/NON-MARKET* analyses of the independent samples presented herein also indicate that with *no reminder*, equality of hypothetical and actual cash WTP is rejected for those in the *MARKET* for art prints and those *NOT IN THE MARKET* at the 0.01 level of significance. With the *reminder*, however, we cannot reject equality of hypothetical and actual cash WTP for those in the *MARKET* or for those *NOT IN THE MARKET* at the 0.05 level of significance. The implication is that hypothetical bias exists for both those in the *MARKET* and those *NOT IN THE MARKET*, but the *reminder* statement was effective in eliminating any significant hypothetical bias for both these groups of people. These findings are confirmed by a comparison of the paired samples.

The LAD regressions indicate that WTP behavior is very different for those in the *MARKET* for art prints and those *NOT IN THE MARKET.* With *no reminder*, B_0=25 and B_1=0 for those in the *MARKET* for art prints, and B_0=0 and B_1=0.09 for those *NOT IN THE MARKET.* With the *reminder*, B_0=-10 and B_1=1 for those in the *MARKET* for art prints, and B_0=0 and B_1=0.17241 for those *NOT IN THE MARKET.*

The difference between the comparisons of the WTP behavior for those in the *MARKET* with the *OTHERS* and those in the *MARKET* with those *NOT IN THE MARKET* is that in the paired sample tests for the equality of medians, we cannot reject equality of hypothetical and actual cash WTP with the *reminder* for those in the *MARKET* but we do reject equality for the *OTHERS*; whereas in the second comparison, we cannot reject equality of hypothetical and actual cash WTP with the *reminder* for those in the *MARKET* or for those *NOT IN THE MARKET.* The implication is that

although the *reminder* was effective in eliminating hypothetical bias for those in the *MARKET* for art prints and those *NOT IN THE MARKET FOR ART PRINTS*, it was not effective in eliminating hypothetical bias for those who neither agreed nor disagreed that they were in the *MARKET* for art prints.

Lastly, the analyses presented herein further indicate that mean or median hypothetical/actual cash WTP ratios frequently reported as measures of hypothetical bias may not be reliable indicators of hypothetical bias. Indeed, in the paired samples, equality of *WTP(h:reminder)* and *WTP(a)* could not be rejected even though the median hypothetical/actual cash WTP ratio for this sample was 7.5:1. Also in the paired samples, equality of *WTP(h:reminder)* and *WTP(a)* was not rejected for those in the *MARKET* but it was rejected for the *OTHERS*, even though the median hypothetical/actual cash WTP ratio was 2.33 for those in the *MARKET* and 2.00 for the *OTHERS*. Therefore, care should be taken in using this ratio as a measure of hypothetical bias.

CHAPTER SIX

CONCLUSION

SUMMARY AND CONCLUSIONS

In summary, analyses of the open ended data presented in Chapter Two indicated that with *no reminder* hypothetical willingness to pay (WTP) was significantly greater than actual cash WTP. With the *reminder*, however, there was mixed evidence with regard to whether there was a significant difference between hypothetical and actual WTP. The *reminder* succeeded in reducing the difference from a ratio of about 3:1 to a ratio of about 1.8:1.

In Chapter Three, analyses of the raw dichotomous choice (DC) data with the Multi-Response Permutation Procedures and Fisher's Exact Test indicated that the percentage of respondents who responded NO (YES) to the posited bid amounts were not significantly different in the hypothetical and actual cash treatments. Analyses of the estimated logit data with the Mann-Whitney test of medians, however, indicated that hypothetical and actual cash WTP were significantly different. However, the log likelihood ratio test of equality of the logit coefficients for both model specifications and the method of convolutions all supported rejecting equality at the .05 level of significance but not at the .01 level, and the confidence interval approach indicated that the estimates of hypothetical and actual cash WTP overlap in the tails. Thus, in spite of the fact that normality is imposed on the logit estimates, even they supported not rejecting equality of hypothetical and actual cash WTP with the DC question format. It was recommended in Chapter Three that the least absolute deviation regression (of medians) be used instead of ordinary least squares in the logit estimation of WTP from DC data because it is problematic that the methodology used in Chapter Three to estimate WTP from DC data imposes normality and continuity on otherwise non-normal and non-continuous data. The

analyses presented in Chapter Three further indicated that there was no starting point bias in either the hypothetical or actual cash treatments.

Comparison of the DC logit estimates of WTP with the paired comparison (PC) measures of willingness to accept (WTA) in Chapter Four indicated that WTA was significantly greater than WTP in spite of the fact that the chooser reference point employed in the experiment was expected to lead to WTP=WTA. Thus the experimental results do not support Kahnemann et al.'s (1990) proposal that endowment effects and loss aversion explain the disparity between WTA and WTP. The chapter concluded that the method of PC (as practiced in Chapter Four) yields invalid estimates of WTA because the method violates microeconomics' fundamental axioms of rational choice, most significantly by excluding the option of indifference between the two goods or one good and a sum of money presented to the respondents, which option is necessary for the completeness axiom of rational choice. The axiom of completeness is a necessary condition for the other two axioms of rational choice, namely transitivity and continuity. Chapter Four lastly proposed as an explanation for the disparities between WTP and WTA in private goods experiments the fact that respondents are not likely to be in the market for the private goods used in laboratory experiments and the fact that WTA respondents have consistently received a free good. The conclusion cites a Working Paper (Lucero), which will report on the findings of an experiment which eliminated both of these experimental design flaws and found that when it can be ascertained that respondents are in the market for the private experimental good and no one receives a free good, then WTP=WTA for relatively inexpensive private goods.

Chapter Five presented my independent analyses of the effect of respondents' being in the *MARKET* or not being in the *MARKET* for private goods used in laboratory experiments on WTP and hypothetical bias. The *MARKET/NON-MARKET* analyses of the independent samples indicated that *WTP(h:reminder)*, *WTP(h:no reminder)*, and *WTP(a)* were all significantly greater for those in the *MARKET* for art prints than for those *NOT IN THE MARKET* at the 0.05 level of significance. The significance of this finding is reflected in the fact that this is the only comparison in which equality of one measure of actual cash WTP with another measure of actual cash WTP can be rejected. Thus those in the *MARKET* for art prints and those *NOT IN THE MARKET* come from two separate and distinct populations with separate and distinct WTP behaviors. Therefore, it would be statistically inappropriate to combine them in one sample. It is the WTP for those *NOT IN THE MARKET*, which is characterized by many zero and low bids, that is responsible for the skewness reflected in our original independent samples.

The implication of these findings for the applicability of criterion validity tests which utilize private goods in laboratory experiments to the contingent valuation method (CVM) is that only WTP measures from those in the *MARKET* for a private experimental good are valid observations. WTP measures from respondents who are *NOT IN THE MARKET* for a private experimental good are simply not meaningful. The inconsistencies between the findings of the independent and paired samples presented in this chapter also raise the issue of whether independent or paired samples are more appropriate in testing for the presence hypothetical bias. Is it the equality of one person's hypothetical and actual cash WTP or the equality of one person's hypothetical WTP with another person's actual cash WTP that is relevant in testing for hypothetical bias?

Independent comparisons can be very misleading if the samples are not characterized by the same percentage of respondents in the *MARKET* or *NOT IN THE MARKET* for any private good used in any laboratory experiment. For example, if the hypothetical WTP treatment consisted mostly of respondents who were in the *MARKET* and the actual WTP treatment consisted mostly of respondents *NOT IN THE MARKET*, then any actual hypothetical bias present would be exaggerated. On the contrary, if the situation were reversed, then any actual hypothetical bias present would be underestimated, and we could end up with results where actual cash WTP is actually less than hypothetical WTP, as suggested by the .857 measure of hypothetical bias reflected in Table 5.9. Although tests for differences in the percentage of NO (YES) responses in the hypothetical and actual cash dichotomous choice treatments between those in the *MARKET* and those *NOT IN THE MARKET* for art prints were not run, a visual inspection of contingency Tables 5.13 – 5.16 clearly indicates that the dichotomous choice data confirms the findings from comparisons of the open-ended data that WTP behavior is significantly different for those in the *MARKET* and those *NOT IN THE MARKET*.

The *MARKET/NON-MARKET* analyses of the independent samples presented herein also indicate that with *no reminder*, equality of hypothetical and actual cash WTP is rejected for those in the *MARKET* for art prints and those *NOT IN THE MARKET* at the 0.01 level of significance. With the *reminder*, however, we cannot reject equality of hypothetical and actual cash WTP for those in the *MARKET* or for those *NOT IN THE MARKET* at the 0.05 level of significance. The implication is that hypothetical bias exists for both those in the *MARKET* and those *NOT IN THE MARKET*. The two highest bids

for $150 and $400 in the hypothetical treatment with *no reminder* and the highest bid for $100 in the hypothetical treatment with the *reminder* were all submitted by respondents who reported that they were in the *MARKET* for art prints. However, the *reminder* statement was effective in eliminating any significant hypothetical bias for both of these populations. These findings are confirmed by a comparison of the paired samples.

The LAD regressions indicate that WTP behavior is very different for those in the *MARKET* for art prints and those *NOT IN THE MARKET*. With *no reminder*, B_0=25 and B_1=0 for those in the *MARKET* for art prints, and B_0=0 and B_1=0.09 for those *NOT IN THE MARKET*. With the *reminder*, B_0=-10 and B_1=1 for those in the *MARKET* for art prints, and B_0=0 and B_1=0.17241 for those *NOT IN THE MARKET*.

The difference between the comparisons of the WTP behavior for those in the *MARKET* with the *OTHERS* and those in the *MARKET* with those *NOT IN THE MARKET* is that in the paired sample tests for the equality of medians, we cannot reject equality of hypothetical and actual cash WTP with the *reminder* for those in the *MARKET* but we do reject equality for the *OTHERS*; whereas in the second comparison, we cannot reject equality of hypothetical and actual cash WTP with the *reminder* for those in the *MARKET* or for those *NOT IN THE MARKET*. The implication is that although the *reminder* was effective in eliminating hypothetical bias for those in the *MARKET* for art prints and those *NOT IN THE MARKET FOR ART PRINTS*, it was not effective in eliminating hypothetical bias for those who neither agreed nor disagreed that they were in the *MARKET* for art prints.

Lastly, the analyses presented herein further indicate that mean or median hypothetical/actual cash WTP ratios frequently reported as measures of hypothetical bias

may not be reliable indicators of hypothetical bias. Indeed, in the paired samples, equality of *WTP(h:reminder)* and *WTP(a)* could not be rejected for those in the *NOT IN THE MARKET*, even though the median hypothetical/actual cash WTP ratio for this sample was 7.5:1. Also in the paired samples, equality of *WTP(h:reminder)* and *WTP(a)* was not rejected for those in the *MARKET* but it was rejected for the *OTHERS*, even though the median hypothetical/actual cash WTP ratio was 2.33 for those in the *MARKET* and 2.00 for the *OTHERS*. Therefore, care should be taken in using this ratio as a measure of hypothetical bias.

There is clearly great danger in applying the findings from laboratory experiments utilizing private goods to the analysis of CVM for public goods, for the problem there is not that individuals are not in the market for public goods, but rather that there is no market for public goods. Therefore, the problems of hypothetical bias and WTP/WTA disparities cannot possibly have the same source in private goods experiments as they do in CVM valuation of public goods. CVM simply assumes away the problem of market failure for public goods and so may not even have face validity. The differences in market characteristics for private goods and public goods are critical, and the sources of hypothetical bias in private goods experiments and public goods CVM must be very different. A valid measure of the tradeoffs involved in allocating life-sustaining environmental resources for one use over another will take into account these differences and not just assume them away, for private and public goods have very separate and distinct characteristics, and payment for their consumption is allocated very separately and distinctly by respondents. This distinction must be taken into consideration by

economists. We cannot simply impose a market upon public goods in the form of CVM as it is currently practiced and expect it to operate as a market for a private good.

REFERENCES

Alberini, Anna. 1995. "Efficiency vs Bias of Willingness-to-Pay Estimates: Bivariate and Interval-Data Models." *Journal of Environmental Economics and Management* (29):169-80.

Arrow, Kenneth, Robert Solow, Paul Portney, Edward Leamer, Roy Radner, and Howard Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." *Federal Register* (58)(10):4602-14.

Bailey, Kenneth. 1987. *Methods of Social Research*. New York: Collier Macmillian Publishers.

Bishop, Richard, and Thomas Heberlein. 1979. "Measuring Values of Extra-Market Goods: Are Indirect Measures of Value Biased?" *American Journal of Agricultural Economics* 61(Dec.):926-30.

Bishop, Richard, Michael Welsh, and Thomas. Heberlein. 1992. *Some Experimental Evidence on the Validity of Contingent Valuation*. Department of Agricultural Economics, University of Wisconsin.

Blalock, Hubert. 1972. *Social Statistics*. 2d ed. New York: McGraw Hill.

Boyce, Rebecca, Thomas Brown, Gary McClelland, George Peterson, and William Schulze. 1992. "An Experimental Examination of Intrinsic Values as a Source of WTA-WTP Disparity." *American Economic Review* 82(Dec.):1366-73.

Brown, Thomas, Patricia Champ, Richard Bishop, and Daniel McCollum. 1996. "Which Response Format Reveals the Truth About Donations to a Public Good." *Land Economics* 72(2):152-66.

Cade, Brian S. and Jon D. Richards. August 2001. *User Manual for Blossom Statistical Software*. U.S. Geological Survey: Midcontinent Ecological Science Center.

Cameron, Trudy. 1988. "A New Paradigm for Valuing Non-Market goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression." *Journal of Environmental Economics and Management* (15):355-80.

Cameron, Trudy and John Quiggin. 1994. "Estimation Using Contingent Valuation Data from a "Dichotomous Choice with Follow-up" Questionnaire." *Journal of Environmental Economics and Management* (27):218-34.

Carson, R. T., J. L Wright, N. J. Carson, A. Alberini, and N. E. Flores. 1994. "A Bibliography of Contingent Valuation Studies and Papers." Natural Resource Damage Assessment Inc., La Jolla, CA.

Champ, Patricia, Richard Bishop, Thomas Brown and Daniel McCollum. 1994. "Some Evidence Concerning the Validity of Contingent Valuation: Preliminary Results of an Experiment." In *Benefits and Costs Transfers in Natural Resources Planning*, compiler R. Ready. Department of Agricultural Economics, University of Kentucky.

Cummings, Ronald G., David S. Brookshire, and William D. Schulze, eds. 1986. "Valuing Environmental Goods: A State of the Arts Assessment of the Contingent Method." Totowa, NJ: Rowman and Allanheld.

Cummings, Ronald, Glenn Harrison, and E. E. Rutstrom. 1995. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?" *American Economic Review* 85(Mar.):260-66.

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

Davis, D. and C. Holt. 1993. *Experimental Economics.* Princeton University Press.

Department of Interior. 1994. "Natural Resource Damage Assessments: Proposed Rule." *Federal Register* 43CFR (May 4) 59(85):23098-111.

Diamond, P. and J. Hausman. 1994. "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives* 8(4):45-64.

Duffield, John, and David Patterson. 1991 "Inference and Optimal Design for a Welfare Measure in Dichotomous Choice Contingent Valuation." *Land Economics* 67(2):225-239.

Duffield, John and David Patterson. 1992. "Field Testing Existence Values: Comparison of Hypothetical and Cash Transaction Values." In *Benefits and Costs in Natural Resource Planning*, Fifth Interim Report, compiler, B. Rettig. Department of Agricultural and Resource Economics, Oregon State University.

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

Fisher, Ann, Gary H. McClelland, and William D. Schulze. 1988. "Measures of Willingness to Pay Versus Willingness to Accept: Evidence, Explanations, and Potential Reconciliation." In *Amenity Resource Valuation: Integrating Economics with Other Disciplines*, eds. George L. Peterson, B. L. Driver, and Robin Gregory, 127-134. State College, PA: Venture Publishing, Inc.

Franciosi, Robert, Praveen Kujai, Roland Michelitsch, and Vernon Smith. 1993. *Experimental Tests of the Endowment Effect.* Economic Science Laboratory, University of Arizona.

Gibbons, Jean Dickinson. 1993. *Nonparametric Statistics, An Introduction.* CA: Sage Publications.

Greene, William. 1990. *Econometric Analysis.* New York: Macmillian, Inc.

Gujarati, Damodar N. 1995. "Regression on Dummy Dependent Variable: The LMP, Logit, Probit, and Tobit Models." In *Basic Econometrics* 540-583. 3rd ed. West Point: McGraw-Hill, Inc.

Hall, Robert, Jack Johnston, and David Lilien. 1990. *MicroTSP User Manual.* CA: Quantitative Micro Software.

Hanemann, Michael. 1991. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *American Economic Review* 635-47.

Hanemann, Michael. 1984. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." *American Journal of Agricultural Economics* 66(3):332-41.

Hanemann, Michael. 1989. "With Discrete Response Data: Reply." *American Journal of Agricultural Economics* 71(4):1057-61.

Hanemann, Michael, John Loomis, and Barbara Kanninen. 1991. "Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation." *American Journal of Agricultural Economics* (73):1255-63.

Hoehn, John and John Loomis. 1993. "Substitution Effects in the Contingent Valuation of Multiple Environmental Programs." *Journal of Environmental Economics and Management* (25):56-75.

Hoehn, John and Alan Randall. 1987. "A Satisfactory Benefit-Cost Indicator for Contingent Valuation." *Journal of Environmental Economics and Management* 14:226-47.

Horowitz, John, Kenneth McConnell, and John Quiggin. 1996. "A Test of Competing Explanations of Compensation Demanded." Unpublished paper, Department of Agricultural Economics, University of Maryland, College Park, MD.

Kahnemann, Daniel, Jack Knetsch, and Richard Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy* (98): 1325-48.

Kanninen, B. 1995. "Bias in Discrete Response Contingent Valuation." *Journal of Environmental Economics and Management* 28(1):114-25.

Kealy, Mary Jo, John Dovidio, and Mark Rockel. 1988. "Accuracy in Valuation is a Matter of Degree." *Land Economics* 64(May):158-71.

Kealy, M. J. and R. W. Turner. 1993. "A Test of the Equality of Close-ended and Open-ended Contingent Valuations." *American Journal of Agricultural Economics* 75:321-31.

Kendall, M. and J. Gibbons. 1990. *Rank Correlation Methods*. 5th ed. London: Edward Arnold, Hodder, and Stoughton Limited.

Kennedy, Peter. 1992. *A Guide to Econometrics*. 3rd ed. Cambridge: MIT Press.

Kmenta, J. 1986. *Elements of Econometrics*. 2d ed. New York: Macmillian, Inc..

Kristrom, B. 1990. "A Non-Parametric Approach to the Estimation of Welfare Measures in Discrete Choice Response Valuation Studies." *Land Economics* 66(2):135-39.

Loomis, John, Thomas Brown, Beatrice Lucero, and George Peterson. 1996. "Improving Validity Experiments of Contingent Valuation Methods: Results of Efforts to Reduce the Disparity of Hypothetical and Actual WTP." *Land Economics* 72(4):450-61.

Loomis, John, Thomas Brown, Beatrice Lucero, and George Peterson. 1997. "Evaluating the Validity of the Dichotomous Choice Question Format in Contingent Valuation." *Environmental and Resource Economics* 10:109-23.

Loomis, John, G. Peterson, P. Champ, T. Brown, and B. Lucero. 1998. "Paired Comparison Estimates of Willingness to Accept Versus Contingent Valuation Estimates of Willingness to Pay." *Journal of Economic Behavior and Organization* 35:501-15.

Lucero, Beatrice. (Working Paper.) "A Reanalysis of the Empirical Evidence Regarding Willingness to Pay and Willingness to Accept in Private Goods Experiments."

McConnell, Kenneth. 1990. "Models for Referendum Data: The Structure of Discrete Choice Models for Contingent Valuation." *Journal of Environmental Economics and Management* 18(1):19-34.

Mielke, P.W. and J.J. Berry. 2001. *Permutation Methods: A Distance Function Approach*. Springer-Verlag.

Mielke, P.W. May 3, 2003. Personal e-mail Communication.

Mielke, Paul, Jr. 1984. "Meteorological Applications of Permutation Techniques based on Distance Functions." In *Handbook of Statistics*, vol. 4, eds. P.R. Krishnaiah and P.K. Sen, 813-30. New York: Elsevier Science Publishers.

Mielke, Paul W., Jr. and Kenneth J. Berry. 1985. "Non-Asymptotic Inferences Based on the Chi-Square Statistic for r by c Contingency Tables." *Journal of Statistical Planning and Inference* 12:41-45.

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

National Oceanic and Atmospheric Administration. 1994. "Natural Resource Damage Assessments: Proposed Rule." *Federal Register* 59CFR(January 7):1062.

Neill, Helen, Ronald Cummings, Philip Ganderton, Glenn Harrison, and Thomas McGuckin. 1994. "Hypothetical Surveys and Real Economic Commitments." *Land Economics* 70(May):145-54.

Nicholson, Walter. 1992. "Preferences and Utility." In *Microeconomic Theory: Basic Principles and Extensions* 73-99. 5th ed. Amherst College: The Dryden Press.

Park, Tom, John Loomis, and Michael Creel. 1991. "Confidence Intervals for Evaluating Benefit Estimates from Dichotomous Choice Contingent Valuation Studies." *Land Economics* 67(1):64-73.

Poe, G., E. Severance-Lossin, and M. Welsh. 1994. "Measuring the Differences in Simulated Distributions: A Convolutions Approach." *American Journal of Agricultural Economics* 76(4):904-15.

Randall, Alan and John R. Stoll. 1980. "Consumer's Surplus in Commodity Space." *American Economic Review*. 70(3):449-455.

Seip, K. and J. Strand. 1992. "Willingness to Pay for Environmental Goods in Norway: A Contingent Valuation Study with Real Payments." *Environmental and Resource Economics* 2:91-106.

Slauson, William L., Brian S. Cade, and Jon D. Richards. 1991. *User Manual for BLOSSOM Statistical Software.* National Ecology Research Center, U.S. Fish and Wildlife Service.

Snedecor, G. and W. Cochran. 1980. *Statistical Methods*. 7th ed. Ames, IA: Iowa State University Press.

Thurston, L. 1927. "A Law of Comparative Judgement." *Psychological Review* (34):273-86.

Welsh, Michael. 1986. "Exploring the Accuracy of the Contingent Valuation Method: Comparison with Simulated Markets." Ph.D. diss., University of Wisconsin, Madison.

Welsh, Michael and Richard Bishop. 1993. "Multiple Bounded Discrete Choice Models." In *Benefits and Costs in Natural Resource Planning*, Compiler J. Bergstrom. Department of Agricultural and Applied Economics, University of Georgia, Athens.

Willig, Robert D. 1976. "Consumer's Surplus Without Apology." *American Economic Review* 66(4):589-597.